

# **IBM SPSS Complex Samples 23**



**注释**

使用本信息及其支持的产品之前，请阅读第 45 页的『声明』中的信息。

**产品信息**

此版本适用于 IBM SPSS Statistics V23.0.0 及所有后续发布和修订，除非在新版本中另有说明。

# 目录

<b>第 1 章 复杂样本过程简介</b> . . . . .	<b>1</b>	<b>第 8 章 复杂样本比率</b> . . . . .	<b>21</b>
复杂样本的属性 . . . . .	1	复杂样本比率: 统计 . . . . .	21
“复杂样本”过程的使用 . . . . .	1	复杂样本比率: 缺失值 . . . . .	21
计划文件 . . . . .	2	复杂样本: 选项 . . . . .	22
其他参考 . . . . .	2	<b>第 9 章 复杂样本一般线性模型</b> . . . . .	<b>23</b>
<b>第 2 章 从复杂设计抽样</b> . . . . .	<b>3</b>	复杂样本一般线性模型 . . . . .	23
创建新样本计划 . . . . .	3	复杂样本一般线性模型: 统计 . . . . .	24
抽样向导: 设计变量 . . . . .	3	复杂样本假设检验 . . . . .	24
浏览抽样向导的树控件 . . . . .	4	复杂样本一般线性模型: 估算的平均值 . . . . .	25
抽样向导: 抽样方法 . . . . .	4	复杂样本一般线性模型: 保存 . . . . .	25
抽样向导: 样本大小 . . . . .	4	复杂样本一般线性模型: 选项 . . . . .	26
定义不等大小 . . . . .	5	CSGLM 命令附加功能 . . . . .	26
抽样向导: 输出变量 . . . . .	5	<b>第 10 章 复杂样本 Logistic 回归</b> . . . . .	<b>27</b>
抽样向导: 计划摘要 . . . . .	6	复杂样本 Logistic 回归: 参考类别 . . . . .	27
抽样向导: 抽取样本选择选项 . . . . .	6	复杂样本 Logistic 回归: 模型 . . . . .	27
抽样向导: 抽取样本“输出文件” . . . . .	6	复杂样本 Logistic 回归: 统计 . . . . .	28
抽样向导: 完成 . . . . .	6	复杂样本假设检验 . . . . .	29
修改现有样本计划 . . . . .	7	复杂样本 Logistic 回归: 几率比 . . . . .	29
抽样向导: 计划摘要 . . . . .	7	复杂样本 Logistic 回归: 保存 . . . . .	29
运行现有样本计划 . . . . .	7	复杂样本 Logistic 回归: 选项 . . . . .	30
CSPLAN 和 CSSELECT 命令附加功能 . . . . .	7	CSLOGISTIC 命令附加功能 . . . . .	30
<b>第 3 章 准备复杂样本以进行分析</b> . . . . .	<b>9</b>	<b>第 11 章 复杂样本序数回归</b> . . . . .	<b>31</b>
创建新的分析计划 . . . . .	9	复杂样本序数回归: 响应概率 . . . . .	31
分析准备向导: 设计变量 . . . . .	9	复杂样本序数回归: 模型 . . . . .	32
浏览分析向导的树控件 . . . . .	9	复杂样本序数回归: 统计 . . . . .	32
分析准备向导: 估计方法 . . . . .	10	复杂样本假设检验 . . . . .	33
分析准备向导: 大小 . . . . .	10	复杂样本序数回归: 几率比 . . . . .	33
定义不等大小 . . . . .	10	复杂样本序数回归: 保存 . . . . .	34
分析准备向导: 计划摘要 . . . . .	11	复杂样本序数回归: 选项 . . . . .	34
分析准备向导: 完成 . . . . .	11	CSORDINAL 命令附加功能 . . . . .	35
修改现有分析计划 . . . . .	11	<b>第 12 章 复杂样本 Cox 回归</b> . . . . .	<b>37</b>
分析准备向导: 计划摘要 . . . . .	11	定义事件 . . . . .	38
<b>第 4 章 复杂样本计划</b> . . . . .	<b>13</b>	预测变量 . . . . .	38
<b>第 5 章 复杂样本频率</b> . . . . .	<b>15</b>	界定依时预测变量 . . . . .	38
复杂样本频率: 统计 . . . . .	15	子组 . . . . .	39
复杂样本: 缺失值 . . . . .	16	模型 . . . . .	39
复杂样本: 选项 . . . . .	16	统计 . . . . .	40
<b>第 6 章 复杂样本描述</b> . . . . .	<b>17</b>	图 . . . . .	40
复杂样本描述: 统计 . . . . .	17	假设检验 . . . . .	41
复杂样本描述: 缺失值 . . . . .	18	保存 . . . . .	41
复杂样本: 选项 . . . . .	18	导出 . . . . .	42
<b>第 7 章 复杂样本交叉表</b> . . . . .	<b>19</b>	选项 . . . . .	43
复杂样本交叉表格: 统计 . . . . .	19	CSCOXREG 命令的附加功能 . . . . .	43
复杂样本: 缺失值 . . . . .	20	<b>声明</b> . . . . .	<b>45</b>
复杂样本: 选项 . . . . .	20	<b>商标</b> . . . . .	<b>46</b>

索引 . . . . . 49

---

## 第 1 章 复杂样本过程简介

在传统软件包中，一项对分析过程的固有假设是：数据文件中的观察数据代表从关注的群体选取的简单随机样本。这种假设对越来越多的公司不再适用，研究人员发现以更为结构化的方式获取样本既经济高效又很方便。

使用“复杂样本”选项，可以根据一项复杂设计选择样本，并将设计规范融入数据分析中，从而确保结果是有效的。

---

### 复杂样本的属性

复杂样本在很多方面与简单随机样本不同。在简单随机样本中，各抽样单元格是直接从整个总体中采用不放回方式以等概率 (WOR) 随机选择的。相比之下，给定的复杂样本具有以下部分或全部特征：

**分层。** 分层抽样在总体的非重叠子组（即层次）中独立选择样本。例如，层次可以是社会经济组、工作类别、年龄组或种族组。通过分层，可以确保子组的样本大小足够大，提高整个估计值的精确度，并在不同层次使用不同抽样方法。

**聚类。** 聚类抽样需要选择抽样单元格组（即聚类）。例如，聚类可以是学校、医院或地理区域，抽样单元格可以是学生、病人或市民。聚类在多阶段设计和区域（地理）样本中很常见。

**多阶段。** 在多阶段抽样中，应基于聚类选择第一阶段样本。然后，通过从所选聚类抽取子样本创建第二阶段样本。如果第二阶段样本是基于子聚类的，则可以向样本添加第三阶段。例如，在调查的第一阶段，可以抽取城市样本。然后，从所选城市中，可以抽取家庭样本。最后，从所选家庭中，可以对个人进行民意调查。使用抽样和分析准备向导可以在一个设计中指定三个阶段。

**非随机抽样。** 如果随机选择难以实现，则可以系统（以固定间隔）或顺序方式抽取单元格。

**不等选择概率。** 如果抽取的聚类包含的单元格数不相等，可以使用与大小成正比 (PPS) 的概率进行抽样，以使聚类的选择概率与其所含单元格的比例相等。PPS 抽样还可以使用更多一般加权设计来选择单元格。

**无限制抽样。** 无限制抽样以放回方式 (WR) 选择单元格。因此，单个单元格可能多次选入样本中。

**抽样权重。** 抽样权重是在抽取复杂样本时自动计算的，与目标总体中每个抽样单元格代表的“频率”十分一致。因此，根据样本的权重总和可以估计群体大小。复杂样本分析过程需要抽样权重以正确分析复杂样本。请注意：这些权重应该在“复杂样本”选项内使用，而不应通过“加权个案”过程用于其他分析过程，该过程将权重视为个案重复。

---

### “复杂样本”过程的使用

“复杂样本”过程的使用取决于特定需要。用户主要类型为执行如下任务的人员：

- 根据复杂设计计划和执行调查，可能以后再分析样本。调查人员的主要工具是抽样向导。
- 根据复杂设计分析以前获得的样本数据文件。在使用“复杂样本”分析过程之前，可能需要使用分析准备向导。

无论是哪类用户，都需要向“复杂样本”过程提供设计信息。为便于重新使用，这一信息存储在计划文件中。

## 计划文件

计划文件包含复杂抽样规范。计划文件有两种类型：

**抽样计划。** 抽样向导中给定的指定项定义用于抽取复杂样本的样本设计。抽样计划文件包含这些指定项。抽样计划文件还包含一个缺省分析计划，该计划使用适合指定样本设计的估计方法。

**分析计划。** 此计划文件包含“复杂样本”分析过程正确计算复杂样本的方差估计所需的信息。该计划包括样本结构、每个阶段的估计方法和对所需变量（如样本权重）的引用。使用分析准备向导可以创建和编辑分析计划。

在计划文件中保存指定项有几个好处，包括：

- 调查人员可以指定多阶段抽样计划的第一阶段并立即抽取第一阶段单元格，并为第二阶段收集抽样单元格的信息，然后修改抽样计划以包括第二阶段。
- 不能访问抽样计划文件的分析人员可以指定一个分析计划，然后从每个“复杂样本”分析过程引用该计划。
- 大型公用样本的设计人员可以发布抽样计划文件，这样简化了分析人员指令，也使分析人员不再需要每人都指定自己的分析计划。

---

## 其他参考

有关抽样方法的更多信息，请参阅以下内容：

Cochran, W. G. 1977. *Sampling Techniques*, 3rd ed. New York: John Wiley and Sons.

Kish, L. 1965. *Survey Sampling*. New York: John Wiley and Sons.

Kish, L. 1987. *Statistical Design for Research*. New York: John Wiley and Sons.

Murthy, M. N. 1967. *Sampling Theory and Methods*. Calcutta, India: Statistical Publishing Society.

Särndal, C., B. Swensson, and J. Wretman. 1992. *Model Assisted Survey Sampling*. New York: Springer-Verlag.

---

## 第 2 章 从复杂设计抽样

该抽样向导将指导您完成创建、修改或执行抽样计划文件的步骤。在使用向导之前，应构思好定义明确的目标总体、抽样单元格列表和适当的样本设计。

---

### 创建新样本计划

1. 从菜单中选择:

分析 > 复杂样本 > 选择样本...

2. 选择**设计样本**并选择一个计划文件名来保存样本计划。
3. 单击**下一步**使向导继续。
4. 或者，在“设计变量”步骤中，可以定义层次、聚类 and 输入样本权重。定义这些内容之后，单击**下一步**。
5. 或者，在“抽样方法”步骤中，可以选择一个方法用于选择样本。

如果选择 **PPS Brewer** 或 **PPS Murthy**，可以单击**完成**抽取样本。否则，单击**下一步**，然后:

6. 在“样本大小”步骤中，指定要抽样的单元格数或单元格比例。
7. 现在，即可单击**完成**抽取样本。

或者，可以进一步执行以下步骤:

- 选择要保存的输出变量。
- 向设计添加第二或第三阶段。
- 设置各选择选项，包括抽取样本的阶段、随机数种子，以及是否将用户缺失值视为设计变量的有效值。
- 选择输出数据的保存位置。
- 将所选项粘贴为命令语法。

---

### 抽样向导: 设计变量

在这一步骤中，可以选择分层变量和聚类变量，可以定义输入样本权重。还可指定阶段的标签。

**分层依据。** 分层变量的交叉分类定义了不同的子群体，即层次。分别为各层获取了不同的样本。要提高估计值的精确度，层中单元格的特征应尽量均一。

**聚类。** 聚类变量定义观察单元格组，即聚类。如果从总体直接抽取观察单元格很昂贵，或者不可能实现，就可以使用聚类；可以从总体抽取聚类，然后从所选聚类抽取观察单元格。但是，使用聚类会在抽样单元格之间引入相关性，导致精度下降。要使这种影响减到最小，聚类中的单元格的特征应尽量均一。必须至少定义一个聚类变量才能计划多阶段设计。在使用多个不同抽样方法时，聚类也是必不可少的。请参阅第 4 页的『抽样向导: 抽样方法』主题以获取更多信息。

**输入样本权重。** 如果当前样本设计是更大样本设计的一部分，则可以从更大样本设计的以前阶段获得样本权重。在当前设计的第一阶段，可以指定一个包含这些权重的数值型变量。对于当前设计的后续阶段，样本权重将自动计算。

**阶段标签。** 可为每个阶段指定一个可选的字符串标签。该标签用在输出中以帮助识别分阶段信息。

注：源变量列表的内容在所有向导步骤中都相同。换言之，在某个特定步骤中从源列表移去的变量将在所有步骤中从该列表移去。返回源列表的变量在所有步骤中都会显示在列表中。

## 浏览抽样向导的树控件

在抽样向导的每个步骤中，左侧都是所有步骤的概要。单击概要中已启用步骤的名称可浏览该向导。只要前面的所有步骤有效（即前面每个步骤都具有要求的最小指定项），则步骤为启用状态。有关给定步骤无效原因的更多信息，请参阅各步骤的“帮助”。

---

## 抽样向导：抽样方法

在这一步骤中，可以指定从活动数据集中选择个案的方式。

**方法。** 该组中的控件用于选择一种选择方法。某些抽样类型允许选择放回抽样 (WR) 或不放回抽样 (WOR)。有关更多信息，请参阅类型描述。请注意，某些与大小成正比的概率 (PPS) 类型只在定义聚类之后才可用，所有 PPS 类型只在设计的第一阶段才可用。此外，WR 方法只在设计的最后阶段才可用。

- **简单随机抽样。** 以等概率选择单元格。单元格可以采用放回或不放回方式进行选择。
- **简单系统。** 在整个抽样框架或层次（如果指定）中，采用不放回方式以固定间隔选择单元格。在第一个区间内随机选择的单元格即选作起始点。
- **简单顺序。** 采用不放回方式以等概率顺序地选择单元格。
- **PPS。** 这是第一阶段方法，它以与大小成正比的概率随机选择单元格。任何单元格都可以采用放回方式选择；只有聚类可以采用不放回方式抽样。
- **PPS 系统。** 这是第一阶段方法，它以与大小成正比的概率系统地选择单元格。并且单元格是以不放回方式选择的。
- **PPS 顺序。** 这是第一阶段方法，它以与聚类大小成正比的概率采用不放回方式顺序选择单元格。
- **PPS Brewer。** 这是第一阶段方法，它以与聚类大小成正比的概率采用不放回方式从每个层次选择两个聚类。要使用此方法，必须指定聚类变量。
- **PPS Murthy。** 这是第一阶段方法，它以与聚类大小成正比的概率采用不放回方式从每个层次选择两个聚类。要使用此方法，必须指定聚类变量。
- **PPS Sampford。** 这是第一阶段方法，它以与聚类大小成正比的概率从每个层次采用不放回方式选择两个以上聚类。它是 Brewer 方法的扩展。要使用此方法，必须指定聚类变量。
- **在分析中使用 WR 估计。** 缺省情况下，估计方法是在计划文件中指定的，与所选抽样方法一致。这样，即使抽样方法意味着 WOR 估计，也可以使用放回方式估计。此选项只在阶段 1 可用。

**大小测量 (MOS)。** 如果选择 PPS 方法，则必须指定定义每个单元格大小的规模测量。这些规模可以在一个变量中显式定义，也可以根据数据计算。或者，可以设置 MOS 的上限和下限，覆盖所有 MOS 变量中的值或根据数据计算的值。这些选项只在阶段 1 可用。

---

## 抽样向导：样本大小

在这一步骤中，可以指定当前阶段中要抽样的单元格数或单元格比例。样本大小可以是固定的，也可以各层不同。为了指定样本大小，前面阶段中选择的聚类可用于定义层次。

**单位。** 可以指定要抽样的单元格的确切样本大小或比例。

- **值。** 应用于所有层次的单个值。如果将**计数**选作单元格度量，则应输入一个正整数。如果选择**比例**，则应输入一个非负值。除非是放回抽样，否则比例值也应不大于 1。
- **各层不相等的值。** 允许您通过“定义不等大小”对话框逐层输入大小值。



- **从变量中读取值。** 允许您选择包含层次大小值的数值变量。

如果选择比例，则可以设置抽样单元格数的下限和上限。

## 定义不等大小

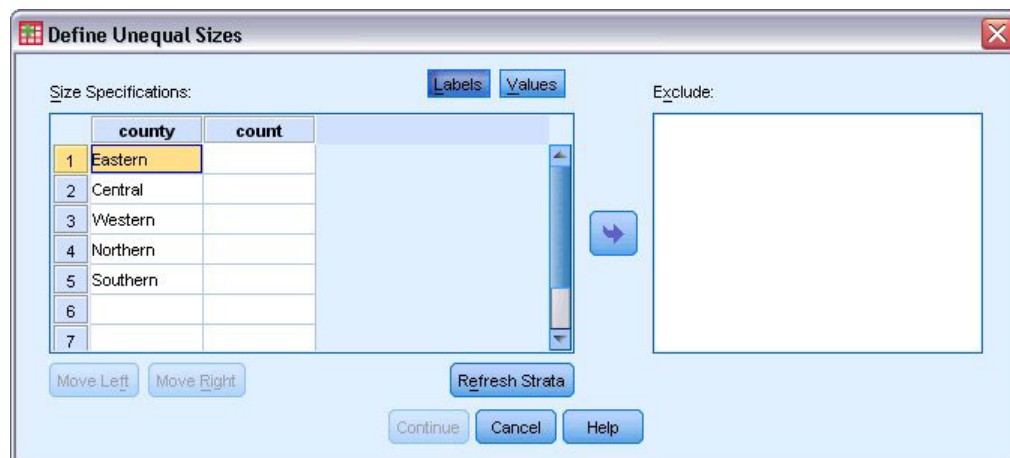


图 1. “定义不等大小”对话框

在“定义不等大小”对话框中，可以逐层输入大小值。

**“大小规范”网格。** 该网格最多显示五个层次变量或聚类变量的交叉分类，即每行一个层次/聚类组合。符合的变量包括当前和以前阶段的所有分层变量，以及以前阶段的所有聚类变量。变量可在网格内重新排序，或者移到“排除”列表。在最右列中输入大小。单击**标签或值**，在网格单元格中分层变量和聚类变量的值标签和数据值的显示之间切换。包含未标注值的单元格始终显示值。单击**刷新层**，用网格中变量的标注数据值的每个组合重新填充网格。

**排除。** 要指定层次/聚类组合子集的大小，请将一个或多个变量移到“排除”列表。这些变量不用于定义样本大小。

## 抽样向导：输出变量

在这一步骤中，可以选择抽取样本时要保存的变量。

**群体大小。** 给定阶段估计的总体单元格数。保存变量的根名为 *PopulationSize\_*。

**样本比例。** 给定阶段的抽样率。保存变量的根名为 *SamplingRate\_*。

**样本大小。** 给定阶段抽取的单元格数。保存变量的根名为 *SampleSize\_*。

**样本权重。** 包含概率的逆。保存变量的根名为 *SampleWeight\_*。

某些分阶段变量是自动生成的。其中包括：

**包含概率。** 给定阶段抽取的单元格比例。保存变量的根名为 *InclusionProbability\_*。

**累积权重。** 当前阶段之前（包括当前阶段）的累积样本权重。保存变量的根名为 *SampleWeightCumulative\_*。

**指标。** 标识给定阶段中多次选择的单元格。保存变量的根名为 *Index\_*。

注：保存变量根名称包含一个整数后缀，该后缀反映阶段号（例如，*PopulationSize\_1\_* 用于阶段 1 保存的群体大小）。

---

## 抽样向导：计划摘要

这是每个阶段的最后一步，提供整个当前阶段的样本设计规范的摘要。在此，可以继续下一阶段（必要时创建阶段），也可以设置抽取样本的选项。

---

## 抽样向导：抽取样本选择选项

在这一步骤中，可以选择是否抽取样本。还可以控制其他抽样选项，如随机种子和缺失值处理。

**抽取样本。** 除了选择是否抽取样本之外，还可以选择执行部分抽样设计。必须按顺序抽取阶段，即抽取阶段 1 之后才能抽取阶段 2。编辑或执行计划时，不能重新抽取锁定的阶段。

**种子。** 它可用于选择生成随机数的种子值。

**包括用户缺失值。** 可确定用户缺失值是否有效。如果有效，则将用户缺失值视为单独的类别。

**数据已排序。** 如果样本框架按照分层变量值预先排序，使用此选项可以加快选择过程。

---

## 抽样向导：抽取样本“输出文件”

在这一步骤中，可以选择抽样个案、权重变量、联合概率和个案选择规则的定向位置。

**样本数据。** 使用这些选项可以确定样本输出的写入位置。样本输出可以添加到活动数据集，也可以写入新数据集，还可以保存到外部 IBM® SPSS® Statistics 数据文件中。数据集在当前会话期间可用，但在后续会话期间不可用，除非显式将其保存为数据文件。数据集名称必须符合变量命名规则。如果指定外部文件或新数据集，则写入所选个案的抽样输出变量和活动数据集中的变量。

**联合概率。** 使用这些选项可以确定联合概率的写入位置。联合概率保存到外部 IBM SPSS Statistics 数据文件。如果选择 PPS WOR、PPS Brewer、PPS Sampford 或 PPS Murthy 方法，并且未指定 WR 估计，则会生成联合概率。

**个案选择规则。** 如果要一次在一个阶段构造样本，可能希望将个案选择规则保存到文本文件中。个案选择规则对于构造后续阶段的子框架非常有用。

---

## 抽样向导：完成

这是最后一步。现在即可保存计划文件并抽取样本，或者将选择内容粘贴到语法窗口中。

在对现有计划文件中的阶段进行更改时，可将已编辑的计划另存为新文件或覆盖现有文件。如果添加阶段而不更改现有阶段，则向导将自动覆盖现有计划文件。如果要将计划保存为新文件，请选择将向导生成的语法粘贴到语法窗口，并在语法命令中更改文件名。

---

## 修改现有样本计划

1. 从菜单中选择:

分析 > 复杂样本 > 选择样本...

2. 选择编辑样本设计，并选择要编辑的计划文件。
3. 单击下一步使向导继续。
4. 在“计划摘要”步骤中复查抽样计划，然后单击下一步。

后续步骤与新设计大体相同。有关更多信息，请参阅各步骤的“帮助”。

5. 浏览到“完成”步骤，为编辑过的计划文件指定新名称，或选择覆盖现有计划文件。

根据需要，您可以:

- 指定已进行抽样的阶段。
- 从计划中移去阶段。

---

## 抽样向导: 计划摘要

在这一步骤中，可以复查抽样计划，确定已进行了抽样的阶段。如果编辑计划，还可以从计划中移去阶段。

**以前抽样的阶段。** 如果扩展抽样框架不可用，则必须一次在一个阶段执行多阶段抽样设计。从下拉列表中选择哪些阶段已进行了抽样。所有执行过的阶段都是锁定的；它们在“抽取样本选择选项”步骤中不可用，并且不能在编辑计划时更改。

**移去阶段。** 可从多阶段设计中移去阶段 2 和 3。

---

## 运行现有样本计划

1. 从菜单中选择:

分析 > 复杂样本 > 选择样本...

2. 选择抽取样本，并选择要运行的计划文件。
3. 单击下一步使向导继续。
4. 在“计划摘要”步骤中复查抽样计划，然后单击下一步。
5. 执行样本计划时，会跳过包含阶段信息的各步骤。现在，随时可以执行“完成”步骤。

根据需要，您可以指定已进行抽样的阶段。

---

## CSPLAN 和 CSSELECT 命令附加功能

使用命令语法语言还可以:

- 为输出变量指定定制名称。
- 在查看器中控制输出。例如，如果设计或修改了样本，则可以取消显示所显示的分阶段计划摘要，如果执行了样本设计，则可以取消显示所显示的按层抽样个案的分布摘要，并请求显示个案处理摘要。
- 在活动数据集中选择一个变量子集，以写入外部样本文件或其他数据集。

请参阅命令语法参考以获取完整的语法信息。



---

## 第 3 章 准备复杂样本以进行分析

分析准备向导将引导您完成创建或修改分析计划的各个步骤，以用于各种“复杂样本”分析过程。使用该向导之前，应先根据一项复杂设计完成样本抽取。

如果不能访问用于抽取样本的抽样计划文件（该抽样计划包含一个缺省分析计划），则创建一个新的计划非常有用。如果确实可以访问用于抽取样本的抽样计划文件，则可以使用抽样计划文件包含的缺省分析计划，也可以覆盖缺省分析指定项并将更改保存到新文件中。

---

### 创建新的分析计划

1. 从菜单中选择:

分析 > 复杂样本 > 准备分析...

2. 选择**创建计划文件**，然后选择一个计划文件名，用于保存分析计划。
3. 单击**下一步**使向导继续。
4. 在“设计变量”步骤中，指定包含样本权重的变量，定义层次和聚类（可选）。
5. 现在，即可单击**完成**保存计划。

或者，可以进一步执行以下步骤:

- 在“估计方法”步骤中，选择用于估计标准误差的方法。
- 在“大小”步骤中，指定抽取的单元格数或每个单元格的包含概率。
- 向设计添加第二或第三阶段。
- 将所选项粘贴为命令语法。

---

### 分析准备向导: 设计变量

在这一步骤中，可以确定分层变量和聚类变量并定义样本权重。还可指定阶段的标签。

**层。** 分层变量的交叉分类定义了不同的子群体，即层次。总样本代表每层的独立样本的组合。

**聚类。** 聚类变量定义观察单元格组，即聚类。多阶段抽取的样本首先在较早阶段中选择聚类，然后从所选聚类中抽取子样本单元格。在分析通过放回方式聚类抽样获得的数据文件时，应将重复指数包括为聚类变量。

**样本权重。** 必须在第一阶段提供样本权重。对于当前设计的后续阶段，样本权重将自动计算。

**阶段标签。** 可为每个阶段指定一个可选的字符串标签。该标签用在输出中以帮助识别分阶段信息。

注: 源变量列表的内容在该向导所有步骤中都相同。换言之，在某个特定步骤中从源列表移去的变量将在所有步骤中从该列表移去。会在所有步骤中显示返回源列表的变量。

### 浏览分析向导的树控件

在每个分析向导步骤的左侧都是所有步骤的概要。单击概要中已启用步骤的名称可浏览该向导。只要前面所有步骤都有效（即只要该步骤前面每个步骤都具有要求的最小指定项），则步骤为启用状态。有关给定步骤无效原因的更多信息，请参阅各步骤的“帮助”。

## 分析准备向导: 估计方法

在这一步骤中，可以指定阶段的估计方法。

**WR (放回式抽样)**。在复杂抽样设计下估计方差时，WR 估计不包括对有限总体抽样的修正 (FPC)。在简单随机抽样 (SRS) 下估计方差时，可以选择包括或排除 FPC。

如果分析权重已进行标度，建议选择不包括用于 SRS 方差估计的 FPC，以免分析权重增加群体大小。SRS 方差估计用于计算类似于设计效果的统计。只能在设计的最后阶段指定 WR 估计；如果选择 WR 估计，向导将不允许添加其他阶段。

**等概率 WOR (等概率不放回式抽样)**。等概率 WOR 估计包括有限总体修正，并假设单元格是等概率抽取的。等概率 WOR 可在设计的任何阶段指定。

**不等概率 WOR (不等概率不放回式抽样)**。除了使用有限总体修正之外，不等概率 WOR 还考虑以不等概率选择的抽样单元格 (通常为聚类)。此估计方法仅在第一阶段可用。

## 分析准备向导: 大小

这一步骤用于指定当前阶段的包含概率或群体大小。大小可以是固定的，也可以各层不同。为了指定大小，前面阶段中指定的聚类可用于定义层次。请注意：仅当选择等概率 WOR 作为估计方法时，这一步骤才是必需的。

**单位**。可以指定精确的群体大小或单元格抽样概率。

- **值**。应用于所有层次的单个值。如果将**群体大小**选作单元格度规，则应输入一个非负整数。如果选择**包含概率**，则应输入一个 0 到 1 之间 (包括 0 和 1) 的值。
- **各层不相等的值**。允许您通过“定义不等大小”对话框逐层输入大小值。
- **从变量中读取值**。允许您选择包含层次大小值的数值变量。

## 定义不等大小

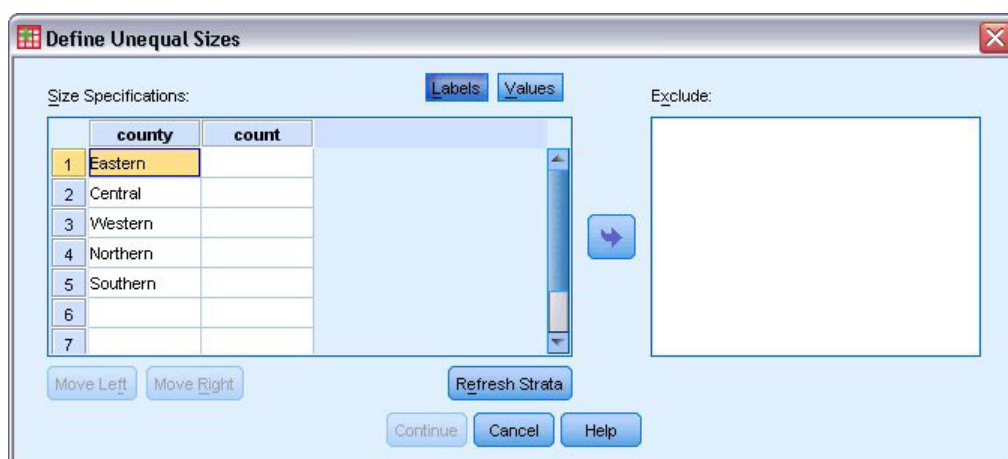


图 2. “定义不等大小”对话框

在“定义不等大小”对话框中，可以逐层输入大小值。

**“大小规范”网格**。该网格最多显示五个层次变量或聚类变量的交叉分类，即每行一个层次/聚类组合。符合的变量包括当前和以前阶段的所有分层变量，以及以前阶段的所有聚类变量。变量可在网格内重新排序，或者移到

“排除”列表。在最右列中输入大小。单击**标签或值**，在网格单元格中分层变量和聚类变量的值标签和数据值的显示之间切换。包含未标注值的单元格始终显示值。单击**刷新层**，用网格中变量的标注数据值的每个组合重新填充网格。

**排除。** 要指定层次/聚类组合子集的大小，请将一个或多个变量移到“排除”列表。这些变量不用于定义样本大小。

---

## 分析准备向导: 计划摘要

这是每个阶段的最后一个步骤，提供整个当前阶段的分析设计规范的摘要。在此，可以继续下一阶段（必要时创建下一阶段），也可以保存分析指定项。

如果不能添加其他阶段，可能的原因如下：

- 在“设计变量”步骤中未指定任何聚类变量。
- 在“估计方法”步骤中选择了 **WR** 估计。
- 这是分析的第三阶段，而该向导最多支持三个阶段。

---

## 分析准备向导: 完成

这是最后一步。现在，可以保存计划文件，或将选择内容粘贴到语法窗口。

在对现有计划文件中的阶段进行更改时，可将已编辑的计划另存为新文件或覆盖现有文件。如果添加阶段而不更改现有阶段，则向导将自动覆盖现有计划文件。如果要将计划保存为新文件，请选择**将本向导生成的语法粘贴到语法窗口**，并在语法命令中更改文件名。

---

## 修改现有分析计划

1. 从菜单中选择：

分析 > 复杂样本 > 准备分析...

2. 选择**编辑计划文件**，然后选择一个计划文件名，用于保存分析计划。
3. 单击**下一步**使向导继续。
4. 在“计划摘要”步骤中复查分析计划，然后单击**下一步**。

后续步骤与新设计大体相同。有关更多信息，请参阅各步骤的“帮助”。

5. 浏览到“完成”步骤，为编辑过的计划文件指定新名称，或选择覆盖现有计划文件。

根据需要，您可以从计划中删除阶段。

---

## 分析准备向导: 计划摘要

在这一步骤中，可以复查分析计划和从计划中移去阶段。

**移去阶段。** 可从多阶段设计中移去阶段 2 和 3。计划必须至少有一个阶段，因此，可以编辑阶段 1，但不能将其从设计中移去。





---

## 第 4 章 复杂样本计划

“复杂样本”分析过程从分析或样本计划文件中获得分析指定项，以便提供有效的结果。

**计划。** 指定分析或样本计划文件的路径。

**联合概率。** 要对用 PPS WOR 方法抽取的聚类使用不等概率 WOR 估计，需要指定包含联合概率的单独文件或打开的数据集。此文件或数据集由抽样向导在抽样过程中创建。



---

## 第 5 章 复杂样本频率

“复杂样本频率”过程可以为所选变量生成频率表并显示单变量统计。您还可以按子组请求统计，子组由一个或多个分类变量定义。

**示例。** 使用“复杂样本频率”过程，基于全美国健康访问调查 (NHIS) 的结果和这一公用数据的适当分析计划，可以获得美国公民维生素使用情况的单变量制表统计。

**统计。** 该过程生成单元格群体大小和表百分比的估计值，以及每个估计值的标准误差、置信区间、变异系数、设计效果、设计效果平方根、累计值和未加权的计数。此外，还计算等单元格比例检验的卡方和似然比统计。

复杂样本频率数据注意事项

**数据。** 要为其生成频率表的变量应为分类变量。子群体变量可以是字符串或数值，但应该是分类变量。

**假设。** 数据文件中的个案代表来自复杂设计的一个样本，该样本应根据在“复杂样本计划”对话框中所选文件内的指定项进行分析。

获取复杂样本频率

1. 从菜单中选择:

    分析 > 复杂样本 > 频率...

2. 选择计划文件。根据需要，选择自定义的联合概率文件。

3. 单击继续。

4. 在“复杂样本频率”对话框中，选择至少一个频率变量。

根据需要，您可以指定变量来定义子群体。统计是针对每个子群体分别计算的。

---

### 复杂样本频率：统计

**单元格。** 在这一组中，可以请求单元格群体大小和表百分比的估计值。

**统计。** 这一组生成与群体大小或表百分比关联的统计。

- **标准误差。** 估计值的标准误差。
- **置信区间。** 估计值的置信区间，使用指定水平。
- **变异系数。** 估计值的标准误差对估计值的比率。
- **去权重计数。** 用于计算估计值的单元格数。
- **设计效果。** 估计值的方差与通过假设样本为简单随机样本所获得的方差的比率。这是指定复杂设计的效果测量，该值与 1 相差越大，表示效果越大。
- **设计效果的平方根。** 这是指定复杂设计的效果测量，该值与 1 相差越大，表示效果越大。
- **累加值。** 通过变量的每个值获得的累计估计值。

**等单元格比例检验。** 对某个类别的变量频率相等的假设生成卡方和似然比检验。对每个变量进行独立检验。

---

## 复杂样本：缺失值

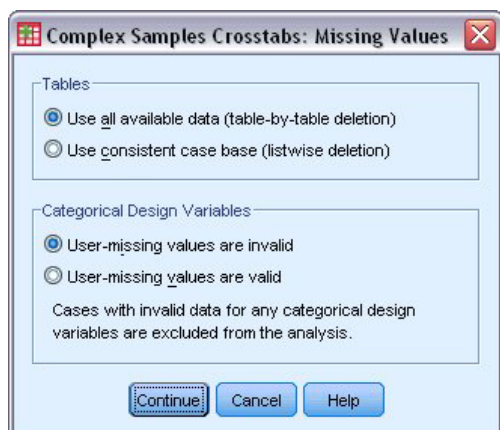


图 3. “缺失值”对话框

表。此组确定在分析中使用的个案。

- **使用所有可用数据。** 缺失值是按照逐个表的方式确定的。因此，在频率表或交叉制表之间，用于计算统计的个案可能不同。
- **确保一致个案基数。** 缺失值是跨所有变量确定的。因此，用于计算统计的个案在所有表中都一致。

类别设计变量。此组确定用户缺失值是否有效。

---

## 复杂样本：选项

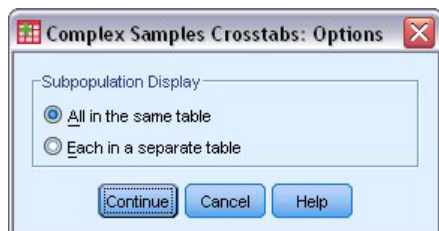


图 4. “选项”对话框

子群体显示。可以选择将子群体显示在同一个表或不同的表中。

---

## 第 6 章 复杂样本描述

“复杂样本描述”过程为多个变量显示单变量汇总统计。您还可以按子组请求统计，子组由一个或多个分类变量定义。

**示例。** 使用“复杂样本描述”过程，基于全美国健康访问调查 (NHIS) 的结果和这一公用数据的适当分析计划，可以获得美国公民活动水平的单变量描述统计。

**统计。** 该过程生成平均值和总和，以及每个估计值的  $t$  检验、标准误差、置信区间、变异系数、未加权的计数、群体大小、设计效果和设计效果平方根。

复杂样本描述数据注意事项

**数据。** 测量应为刻度变量。子群体变量可以是字符串或数值，但应该是分类变量。

**假设。** 数据文件中的个案代表来自复杂设计的一个样本，该样本应根据在“复杂样本计划”对话框中所选文件内的指定项进行分析。

获取复杂样本描述

1. 从菜单中选择:

    分析 > 复杂样本 > 描述...

2. 选择计划文件。根据需要，选择自定义的联合概率文件。

3. 单击继续。

4. 在“复杂样本描述”对话框中，选择至少一个测量变量。

根据需要，您可以指定变量来定义子群体。统计是针对每个子群体分别计算的。

---

### 复杂样本描述: 统计

**摘要。** 在这一组中，可以请求测量变量的平均值和总和的估计值。此外，还可以请求针对指定值进行估计值的  $t$  检验。

**统计。** 这一组生成与平均值或总和关联的统计。

- **标准误差。** 估计值的标准误差。
- **置信区间。** 估计值的置信区间，使用指定水平。
- **变异系数。** 估计值的标准误差对估计值的比率。
- **去权重计数。** 用于计算估计值的单元格数。
- **群体大小。** 总体中估计的单元格数。
- **设计效果。** 估计值的方差与通过假设样本为简单随机样本所获得的方差的比率。这是指定复杂设计的效果测量，该值与 1 相差越大，表示效果越大。
- **设计效果的平方根。** 这是指定复杂设计的效果测量，该值与 1 相差越大，表示效果越大。

---

## 复杂样本描述：缺失值

测量值统计。 此组确定在分析中使用的个案。

- 使用所有可用数据。 缺失值是逐个变量确定的，因此，各测量变量用于计算统计的个案可能不同。
- 确保一致个案基数。 缺失值是通过所有变量确定的，因此，用于计算统计的个案是一致的。

类别设计变量。 此组确定用户缺失值是否有效。

---

## 复杂样本：选项

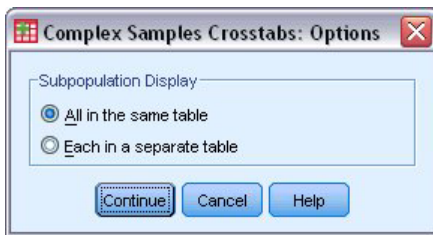


图 5. “选项”对话框

子群体显示。 可以选择将子群体显示在同一个表或不同的表中。

---

## 第 7 章 复杂样本交叉表

复杂样本交叉表格过程可以为所选变量对生成交叉表格并显示二阶统计。您还可以按子组请求统计，子组由一个或多个分类变量定义。

**示例。** 使用“复杂样本交叉表格”过程，基于全美国健康访问调查 (NHIS) 的结果和这一公用数据的适当分析计划，可以获得美国公民维生素使用量和抽烟频率的交叉分类统计。

**统计。** 该过程生成单元格群体大小、行百分比、列百分比和表百分比的估计值，以及每个估计值的标准误差、置信区间、变异系数、期望值、设计效果、设计效果平方根、残差、调整的残差和未加权的计数。计算几率比、相对风险和危险度差值以在 2x2 表中使用。此外，还计算 Pearson 和似然比统计用于行变量和列变量的独立性检验。

复杂样本交叉表格数据注意事项

**数据。** 行变量和列变量应是分类变量。子群体变量可以是字符串或数值，但应该是分类变量。

**假设。** 数据文件中的个案代表来自复杂设计的一个样本，该样本应根据在“复杂样本计划”对话框中所选文件内的指定项进行分析。

获取复杂样本交叉表格

1. 从菜单中选择:

    分析 > 复杂样本 > 交叉表格...

2. 选择计划文件。根据需要，选择自定义的联合概率文件。

3. 单击继续。

4. 找“复杂样本交叉表格”对话框中，选择至少一个行变量和一个列变量。

根据需要，您可以指定变量来定义子群体。统计是针对每个子群体分别计算的。

---

### 复杂样本交叉表格: 统计

**单元格。** 在这一组中，可以请求单元格群体大小、行百分比、列百分比和表百分比的估计值。

**统计。** 这一组生成与群体大小、行百分比、列百分比和表百分比关联的统计。

- **标准误差。** 估计值的标准误差。
- **置信区间。** 估计值的置信区间，使用指定水平。
- **变异系数。** 估计值的标准误差对估计值的比率。
- **期望值。** 在假设行变量和列变量独立的条件下，估计值的期望值。
- **去权重计数。** 用于计算估计值的单元格数。
- **设计效果。** 估计值的方差与通过假设样本为简单随机样本所获得的方差的比率。这是指定复杂设计的效果测量，该值与 1 相差越大，表示效果越大。
- **设计效果的平方根。** 这是指定复杂设计的效果测量，该值与 1 相差越大，表示效果越大。
- **残差。** 如果两个变量之间没有关系，则期望值是期望在单元格中出现的个案数。如果行变量和列变量独立，则正的残差表示单元格中的实际个案数多于期望的个案数。

- **调整的残差。** 单元格的残差（观察值减去期望值）除以其标准误差的估计值。生成的标准化残差表示为平均值上下的标准差单位。

**2x2 表的摘要。** 这一组为每个行变量和列变量都具有两个类别的表生成统计。每一个都是某因子的存在与某事件的发生之间相关性大小的测量。

- **几率比。** 当因子很少出现时，几率比可用作相对风险的估计值。
- **相对危险度。** 存在因子出现事件的风险与不存在因子出现事件的风险的比率。
- **危险度差值。** 存在因子出现事件的风险与不存在因子出现事件的风险之差。

**行和列的独立性检验。** 生成行变量和列变量独立的假设的卡方检验和似然比检验。对每对变量进行单独检验。

---

## 复杂样本：缺失值

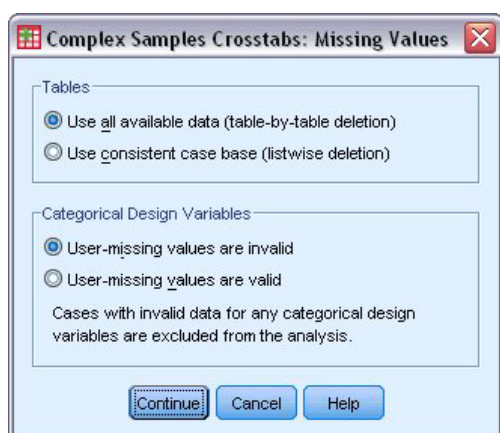


图 6. “缺失值”对话框

**表。** 此组确定在分析中使用的个案。

- **使用所有可用数据。** 缺失值是按照逐个表的方式确定的。因此，在频率表或交叉制表之间，用于计算统计的个案可能不同。
- **确保一致个案基数。** 缺失值是跨所有变量确定的。因此，用于计算统计的个案在所有表中都一致。

**类别设计变量。** 此组确定用户缺失值是否有效。

---

## 复杂样本：选项

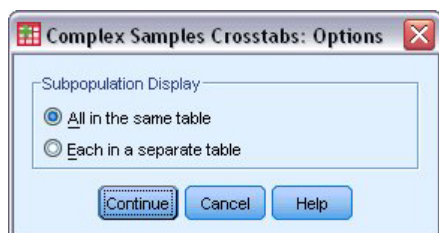


图 7. “选项”对话框

**子群体显示。** 可以选择将子群体显示在同一个表或不同的表中。



---

## 第 8 章 复杂样本比率

“复杂样本比率”过程显示变量的比率的单变量汇总统计。您还可以按子组请求统计，子组由一个或多个分类变量定义。

**示例。** 使用“复杂样本比率”过程，基于全国范围调查（根据一项复杂设计并采用适合数据的分析计划进行）的结果，可以获取当前财产价值与上次评估价值的比率的描述统计。

**统计。** 该过程生成比率估计值、 $t$  检验、标准误差、置信区间、变异系数、未加权的计数、群体大小、设计效果和设计效果平方根。

复杂样本比率数据注意事项

**数据。** 分子和分母应为正值刻度变量。子群体变量可以是字符串或数值，但应该是分类变量。

**假设。** 数据文件中的个案代表来自复杂设计的一个样本，该样本应根据在“复杂样本计划”对话框中所选文件内的指定项进行分析。

获取复杂样本比率

1. 从菜单中选择:

    分析 > 复杂样本 > 比率...

2. 选择计划文件。根据需要，选择自定义的联合概率文件。

3. 单击继续。

4. 在“复杂样本比率”对话框中，选择至少一个分子变量和一个分母变量。

根据需要，您可以指定变量来定义要为其生成统计的子组。

---

### 复杂样本比率: 统计

**统计。** 此组生成与比率估计值关联的统计。

- **标准误差。** 估计值的标准误差。
- **置信区间。** 估计值的置信区间，使用指定水平。
- **变异系数。** 估计值的标准误差对估计值的比率。
- **去权重计数。** 用于计算估计值的单元格数。
- **群体大小。** 总体中估计的单元格数。
- **设计效果。** 估计值的方差与通过假设样本为简单随机样本所获得的方差的比率。这是指定复杂设计的效果测量，该值与 1 相差越大，表示效果越大。
- **设计效果的平方根。** 这是指定复杂设计的效果测量，该值与 1 相差越大，表示效果越大。

**T 检验。** 可以请求针对指定值进行估计值的  $t$  检验。

---

### 复杂样本比率: 缺失值

**比率。** 此组确定在分析中使用的个案。

- 使用所有可用数据。缺失值是按照逐个比率的方式确定的。因此，用于计算统计值的个案在各个分子分母对间可能不同。
- 确保一致个案基数。缺失值是跨所有变量确定的。因此，用于计算统计的个案是一致的。

类别设计变量。此组确定用户缺失值是否有效。

---

## 复杂样本：选项

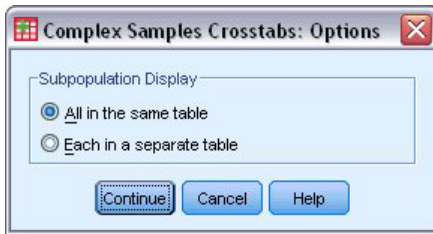


图 8. “选项”对话框

子群体显示。可以选择将子群体显示在同一个表或不同的表中。

---

## 第 9 章 复杂样本一般线性模型

“复杂样本一般线性模型”(CSGLM) 过程对通过复杂抽样方法抽取的样本执行线性回归分析以及方差和协方差分析。您还可以请求对子群体进行分析。

**示例。** 根据一项复杂设计，杂货连锁店对一组顾客的购物习惯进行调查。在获得了调查结果以及每个顾客在上个月的消费金额之后，商店希望了解顾客购物的频率是否与他们在一个月中的消费金额有关，从而针对顾客性别进行控制并采用抽样设计。

**统计。** 该过程生成模型参数的估计值、标准误差、置信区间、 $t$  检验、设计效果和设计效果平方根，以及参数估计值之间的相关系数和协方差；还可以生成模型拟合度的测量和自变量、因变量的描述统计。此外，您还可以请求模型因子和因子交互的水平估计边际平均值。

复杂样本一般线性模型数据注意事项

**数据。** 因变量是定量变量。因子是分类变量。协变量是与因变量相关的定量变量。子群体变量可以是字符串或数值，但应该是分类变量。

**假设。** 数据文件中的个案代表来自复杂设计的一个样本，该样本应根据在“复杂样本计划”对话框中所选文件内的指定项进行分析。

获取复杂样本一般线性模型

1. 从菜单中选择:

分析 > 复杂样本 > 一般线性模型...

2. 选择计划文件。根据需要，选择自定义的联合概率文件。

3. 单击**继续**。

4. 在“复杂样本一般线性模型”对话框中，选择一个因变量。

根据需要，您可以:

- 为因子和协变量选择适合您数据的变量。
- 指定用于定义子群体的变量。仅对子群体变量的所选类别执行该分析。

---

### 复杂样本一般线性模型

**指定模型效应。** 缺省情况下，该过程使用主对话框中指定的因子和协变量构建主效应模型。此外，还可以构建包含交互效应和嵌套项的自定义模型。

非嵌套项

对于选定因子和协变量:

**交互。** 为所有选定变量创建最高级交互项。

**主效应。** 为每个选定的变量创建主效应项。

**所有二阶。** 创建选定变量的所有可能的双向交互。

所有三阶。 创建选定变量的所有可能的三阶交互。

所有四阶。 创建选定变量的所有可能的四阶交互。

所有五阶。 创建选定变量的所有可能的五阶交互。

嵌套项

在此过程中，可为您的模型建立嵌套项。嵌套项有助于对其值不与另一个因子的水平交互作用的因子或协变量的效应进行建模。例如，杂货连锁店可能在不同商店位置迎合顾客的不同消费习惯。由于每位顾客只经常光顾某一位置的商店，因此可以说客户效应嵌套在商店位置效应中。

此外，还可以包含交互效应，例如包含相同协变量的多项式项，或将多层嵌套添加到嵌套项。

限制。 嵌套项有以下限制：

- 一次交互内的所有因子必须是唯一的。因此，如果  $A$  是因子，则指定  $A*A$  是无效的。
- 嵌套效应内的所有因子必须是唯一的。因此，如果  $A$  是因子，则指定  $A(A)$  是无效的。
- 效应不可嵌套在协变量中。因此，如果  $A$  是因子且  $X$  是协变量，则指定  $A(X)$  是无效的。

截距。 模型中通常包含截距。如果您可以假设数据穿过原点，则可以排除截距。即使在模型中包含截距，也可以选择取消显示与之相关的统计。

---

## 复杂样本一般线性模型：统计

模型参数。 使用此组可以控制与模型参数有关的统计的显示。

- 估算。 显示系数的估计值。
- 标准误差。 显示每个系数估计值的标准误差。
- 置信区间。 显示每个系数估计值的置信区间。在“选项”对话框中设置该区间的置信度。
- T 检验。 显示每个系数估计值的  $t$  检验。每个检验的原假设是该系数的值为 0。
- 参数估值协方差。 显示模型系数的协方差矩阵的估计值。
- 参数估值的相关性。 显示模型系数的相关性矩阵的估计值。
- 设计效果。 估计值的方差与通过假设样本为简单随机样本所获得的方差的比率。这是指定复杂设计的效果测量，该值与 1 相差越大，表示效果越大。
- 设计效果的平方根。 这是指定复杂设计的效果测量，该值与 1 相差越大，表示效果越大。

模型拟合度。 显示  $R^2$  和根均方误差统计。

因变量和协变量的群体平均值。 显示有关因变量、协变量和因子的摘要信息。

样本设计信息。 显示有关样本的摘要信息，包括未加权的计数和群体大小。

---

## 复杂样本假设检验

检验统计。 在这一组中，可以选择用于检验假设的统计类型。可以在  $F$ 、调整的  $F$ 、卡方和调整的卡方之间选择。

样本自由度。 在这一组中，可以控制用于计算所有检验统计的  $p$  值的抽样设计自由度。如果基于抽样设计，该值为抽样第一阶段的主抽样单元格数和层数之差。或者，也可以通过指定一个正整数设置自定义自由度。

调整的多重比较。在执行包含多重比较的假设检验时，总体显著性水平可从所包含的的比较的显著性水平进行调节。使用此组可以选择调节方法。

- **显著性最低的差异。** 此方法并不控制拒绝某些线性对比不同于原假设值这一假设的总体概率。
- **连续 Sidak (Sequential Sidak).** 这是一个逐步下降的排斥性 Sidak 过程，就排斥单个假设而言，其保守性小得多，且保持了相同的总体显著性水平。
- **连续 Bonferroni (Sequential Bonferroni).** 这是按顺序逐步降低的拒绝 Bonferroni 过程，在拒绝个别假设方面不保守，但维持相同的总体显著性水平。
- **Sidak.** 此方法提供比 Bonferroni 方法更严格的界限。
- **Bonferroni.** 此方法根据多项对比检验调整观察到的显著性水平。

---

## 复杂样本一般线性模型：估算的平均值

在“估算的平均值”对话框中，可以控制在“模型”子对话框中指定的因子和因子交互水平的模型估计边际平均值的显示。还可以请求显示整个总体的平均值。

项。计算所选因子和因子交互的估计平均值。

对比。对比决定如何设定假设检验以比较估计平均值。

- **简单。** 将每个水平的平均值与指定水平的平均值进行比较。当存在控制组时，此类对比很有用。
- **偏移。** 将每个水平（参考类别除外）的平均值与所有水平的平均值（总平均值）进行比较。因子的水平可以为任何顺序。
- **差分。** 将每个水平（第一个除外）的平均值与先前水平的平均值进行比较。有时候将其称为逆 Helmert 对比。
- **Helmert.** 将因子的每个水平的平均值（最后一个水平除外）与后面水平的平均值进行比较。
- **重复。** 将每个水平的平均值（最后一个水平除外）与后一个水平的平均值进行比较。
- **多项式。** 比较线性效应、二次效应、三次效应等等。第一自由度包含跨所有类别的线性效应；第二自由度包含二次效应，依此类推。这些对比常常用来估计多项式趋势。

参考类别。简单对比和偏移对比需要参考类别或与其他因子级别进行比较的因子级别。

---

## 复杂样本一般线性模型：保存

保存变量。在这一组中，可以将模型预测值和残差保存为工作文件中的新变量。

将模型导出为 **IBM SPSS Statistics 数据**。写入一个 IBM SPSS Statistics 格式的数据集，包含具有参数估计值、标准误差、显著性值和自由度的参数相关性或协方差矩阵。矩阵文件中变量顺序如下。

- **RowType\_。** 取值（或值标签）为 COV（协方差）、CORR（相关性）、EST（参数估计）、SE（标准误差）、SIG（显著性水平）和 DF（抽样设计自由度）。存在每个模型参数的 COV（或 CORR）行类型的单独个案，以及每个其他行类型的个案。
- **VarName\_。** 对于行类型 COV 或 CORR，取值为 P1、P2、...，对应于所有模型参数的有序列表，值标签对应于在参数估计值表中显示的参数字符串。对于其他行类型，单元格为空。
- **P1, P2, ...** 这些变量对应于所有模型参数的有序列表，值标签对应于在参数估计值表中显示的参数字符串，这些变量根据行类型取值。对于冗余参数，所有协方差设为零；相关性设为系统缺失值；所有参数估计值设为零；并且所有标准误差、显著性水平和残差自由度设为系统缺失值。

注：该文件不能立即用于在其他读取矩阵文件的过程中执行进一步分析，除非这些过程接受在此导出的所有行类型。

将模型导出为 **XML**。将参数估计值和参数协方差矩阵（如果选择）以 XML (PMML) 格式保存。您可以使用该模型文件以应用模型信息到其他数据文件用于评分目的。

---

## 复杂样本一般线性模型：选项

**用户缺失值。** 所有设计变量以及因变量和任何协变量都必须包含有效数据。从分析中删除含有任何这些变量的无效数据的个案。使用这些控制可以确定是否在层次、聚类、子群体和因子变量中将用户缺失值视为有效。

**置信区间。** 这是系数估值和估计边际平均值的置信区间水平。指定大于等于 50 且小于 100 的值。

---

## CSGLM 命令附加功能

使用命令语法语言还可以：

- 指定作用对比作用的线性组合或一个值的自定义检验（使用 CUSTOM 子命令）。
- 在计算估计边际平均值时修正值不同于其平均值的协变量（使用 EMMEANS 子命令）。
- 为多项式对比指定度规（使用 EMMEANS 子命令）。
- 指定用于检查奇异性的容差值（使用 CRITERIA 子命令）。
- 为保存的变量创建用户指定名称（使用 SAVE 子命令）。
- 生成常规可估计函数表（使用 PRINT 子命令）。

请参阅 [命令语法参考](#) 以获取完整的语法信息。

---

## 第 10 章 复杂样本 Logistic 回归

“复杂样本 Logistic 回归”过程对通过复杂抽样方法抽取的样本的二元或多项因变量执行 logistic 回归分析。您还可以请求对子群体进行分析。

**示例。** 根据一项复杂设计，信贷员收集了在几个不同分支机构贷款的客户的过去记录。融入样本设计时，信贷员希望了解客户拖欠的概率是否与年龄、工作经历和信用负债量有关。

**统计。** 该过程生成模型参数的估计值、取幂估值、标准误差、置信区间、 $t$  检验、设计效果和设计效果平方根，以及参数估计值之间的相关系数和协方差。还可以生成因变量和自变量的伪  $R^2$  统计、分类表和描述统计。

复杂样本 Logistic 回归数据注意事项

**数据。** 因变量是分类变量。因子是分类变量。协变量是与因变量相关的定量变量。子群体变量可以是字符串或数值，但应该是分类变量。

**假设。** 数据文件中的个案代表来自复杂设计的一个样本，该样本应根据在“复杂样本计划”对话框中所选文件内的指定项进行分析。

获取复杂样本 Logistic 回归

1. 从菜单中选择:

**分析 > 复杂样本 > Logistic 回归...**

2. 选择计划文件。根据需要，选择自定义的联合概率文件。

3. 单击**继续**。

4. 在“复杂样本 Logistic 回归”对话框中，选择一个因变量。

根据需要，您可以:

- 为因子和协变量选择适合您数据的变量。
- 指定用于定义子群体的变量。仅对子群体变量的所选类别执行该分析。

---

### 复杂样本 Logistic 回归: 参考类别

缺省情况下，“复杂样本 Logistic 回归”过程将最高值类别作为参考类别。在此对话框中，可以将最高值、最低值或自定义类别指定为参考类别。

---

### 复杂样本 Logistic 回归: 模型

**指定模型效应。** 缺省情况下，该过程使用主对话框中指定的因子和协变量构建主效应模型。此外，还可以构建包含交互效应和嵌套项的自定义模型。

非嵌套项

对于选定因子和协变量:

**交互。** 为所有选定变量创建最高级交互项。

**主效应。** 为每个选定的变量创建主效应项。

**所有二阶。** 创建选定变量的所有可能的双向交互。

**所有三阶。** 创建选定变量的所有可能的三阶交互。

**所有四阶。** 创建选定变量的所有可能的四阶交互。

**所有五阶。** 创建选定变量的所有可能的五阶交互。

**嵌套项**

在此过程中，可为您的模型建立嵌套项。嵌套项有助于对其值不与另一个因子的水平交互作用的因子或协变量的效应进行建模。例如，杂货连锁店可能在不同商店位置迎合顾客的不同消费习惯。由于每位顾客只经常光顾某一位置的商店，因此可以说**客户效应嵌套在商店位置效应中**。

此外，还可以包含交互效应，例如包含相同协变量的多项式项，或将多层嵌套添加到嵌套项。

**限制。** 嵌套项有以下限制：

- 一次交互内的所有因子必须是唯一的。因此，如果  $A$  是因子，则指定  $A*A$  是无效的。
- 嵌套效应内的所有因子必须是唯一的。因此，如果  $A$  是因子，则指定  $A(A)$  是无效的。
- 效应不可嵌套在协变量中。因此，如果  $A$  是因子且  $X$  是协变量，则指定  $A(X)$  是无效的。

**截距。** 模型中通常包含截距。如果您可以假设数据穿过原点，则可以排除截距。即使在模型中包含截距，也可以选择取消显示与之相关的统计。

---

## 复杂样本 Logistic 回归：统计

**模型拟合度。** 控制度量总体模型性能的统计的显示。

- **伪  $R^2$  方。** 在 Logistic 回归模型中，没有与线性回归的  $R^2$  完全对应的统计。相反，却有多种试图模拟  $R^2$  统计的属性的测量。
- **分类表。** 按因变量的模型预测类别显示观察类别的制表交叉分类。

**参数。** 使用此组可以控制与模型参数有关的统计的显示。

- **估算。** 显示系数的估计值。
- **取幂估值。** 显示以系数估值为幂的自然对数的底数。当该估值对于统计检验有良好的属性时，取幂估值（即  $\exp(B)$ ）更易于解释。
- **标准误差。** 显示每个系数估计值的标准误差。
- **置信区间。** 显示每个系数估计值的置信区间。在“选项”对话框中设置该区间的置信度。
- **T 检验。** 显示每个系数估计值的  $t$  检验。每个检验的原假设是该系数的值为 0。
- **参数估值协方差。** 显示模型系数的协方差矩阵的估计值。
- **参数估值的相关性。** 显示模型系数的相关性矩阵的估计值。
- **设计效果。** 估计值的方差与通过假设样本为简单随机样本所获得的方差的比率。这是指定复杂设计的效果测量，该值与 1 相差越大，表示效果越大。
- **设计效果的平方根。** 这是指定复杂设计的效果测量，该值与 1 相差越大，表示效果越大。

**模型变量的汇总统计。** 显示有关因变量、协变量和因子的摘要信息。

**样本设计信息。** 显示有关样本的摘要信息，包括未加权的计数和群体大小。



---

## 复杂样本假设检验

**检验统计。** 在这一组中，可以选择用于检验假设的统计类型。可以在  $F$ 、调整的  $F$ 、卡方和调整的卡方之间选择。

**样本自由度。** 在这一组中，可以控制用于计算所有检验统计的  $p$  值的抽样设计自由度。如果基于抽样设计，该值为抽样第一阶段的主抽样单元格数和层数之差。或者，也可以通过指定一个正整数设置自定义自由度。

**调整的多重比较。** 在执行包含多重比较的假设检验时，总体显著性水平可从所包含的比較的显著性水平进行调节。使用此组可以选择调节方法。

- **显著性最低的差异。** 此方法并不控制拒绝某些线性对比不同于原假设值这一假设的总体概率。
- **连续 Sidak (Sequential Sidak).** 这是一个逐步下降的排斥性 Sidak 过程，就排斥单个假设而言，其保守性小得多，且保持了相同的总体显著性水平。
- **连续 Bonferroni (Sequential Bonferroni).** 这是按顺序逐步降低的拒绝 Bonferroni 过程，在拒绝个别假设方面不保守，但维持相同的总体显著性水平。
- **Sidak.** 此方法提供比 Bonferroni 方法更严格的界限。
- **Bonferroni.** 此方法根据多项对比检验调整观察到的显著性水平。

---

## 复杂样本 Logistic 回归: 几率比

使用“几率比”对话框，可以控制指定的因子和协变量的模型估计几率比的显示。计算因变量的每个类别（参考类别除外）的一组单独几率比。

**因子。** 对于所选的每个因子，都显示该因子处于每个类别的几率与处于指定参考类别的几率之比。

**协变量。** 对于所选的每个协变量，都显示协变量平均值加上指定变化单位的几率与处于平均值的几率之比。

在计算因子或协变量的几率比时，该过程将修正所有处于其最高级的其他因子以及所有处于其平均值的其他协变量。如果因子或协变量与模型中的其他预测变量交互作用，则几率比不仅取决于指定变量的变化，而且还取决于与之交互的变量的值。如果指定的协变量在模型中与其自身交互作用（例如  $age*age$ ），则几率比同时取决于协变量的变化和协变量的值。

---

## 复杂样本 Logistic 回归: 保存

**保存变量。** 在这一组中，可以将模型预测类别和预测概率保存为活动数据集中的新变量。

将模型导出为 **IBM SPSS Statistics 数据**。写入一个 IBM SPSS Statistics 格式的数据集，包含具有参数估计值、标准误差、显著性值和自由度的参数相关性或协方差矩阵。矩阵文件中变量顺序如下。

- **RowType\_。** 取值（或值标签）为 COV（协方差）、CORR（相关性）、EST（参数估计）、SE（标准误差）、SIG（显著性水平）和 DF（抽样设计自由度）。存在每个模型参数的 COV（或 CORR）行类型的单独个案，以及每个其他行类型的个案。
- **VarName\_。** 对于行类型 COV 或 CORR，取值为 P1、P2、...，对应于所有模型参数的有序列表，值标签对应于在参数估计值表中显示的参数字符串。对于其他行类型，单元格为空。
- **P1, P2, ...** 这些变量对应于所有模型参数的有序列表，值标签对应于在参数估计值表中显示的参数字符串，这些变量根据行类型取值。对于冗余参数，所有协方差设为零；相关性设为系统缺失值；所有参数估计值设为零；并且所有标准误差、显著性水平和残差自由度设为系统缺失值。

注：该文件不能立即用于在其他读取矩阵文件的过程中执行进一步分析，除非这些过程接受在此导出的所有行类型。

**将模型导出为 XML。** 将参数估计值和参数协方差矩阵（如果选择）以 XML (PMML) 格式保存。您可以使用该模型文件以应用模型信息到其他数据文件用于评分目的。

---

## 复杂样本 Logistic 回归：选项

**估值。** 在这一组中，可以控制模型估计中使用的各种条件。

- **最大迭代次数。** 算法将执行的最大迭代次数。指定一个非负整数。
- **最大折半次数。** 每次迭代时，步长都会减去因子 0.5，直到对数似然估计增加或者达到最大步骤对分。指定一个正整数。
- **根据参数估值更改限制迭代。** 如果选择此项，算法将在参数估计值的绝对或相对变化小于指定值（必须为非负值）的迭代之后停止。
- **根据对数似然估计更改限制迭代。** 如果选择此项，算法将在对数似然估计函数的绝对或相对变化小于指定值（必须为非负值）的迭代之后停止。
- **检查数据点的完整分隔。** 如果选择此项，算法将执行检验以确保参数估计值具有唯一值。当过程可生成一个正确对每个个案进行分类的模型时，将发生分离。
- **显示迭代历史记录。** 从第 0 次迭代（初始估计）开始，在每  $n$  次迭代时显示参数估计值和统计。如果选择打印迭代历史记录，则无论  $n$  值为多少，将总是打印最后一次迭代。

**用户缺失值。** 所有设计变量以及因变量和任何协变量都必须包含有效数据。从分析中删除含有任何这些变量的无效数据的个案。使用这些控制可以确定是否在层次、聚类、子群体和因子变量中将用户缺失值视为有效。

**置信区间。** 这是系数估值、取幂系数估值和几率比的置信区间度。指定大于等于 50 且小于 100 的值。

---

## CSLOGISTIC 命令附加功能

使用命令语法语言还可以：

- 指定作用对比作用的线性组合或一个值的自定义检验（使用 CUSTOM 子命令）。
- 在计算因子和协变量的几率比（使用 ODDS RATIOS 子命令）时修正其他模型变量的值。
- 指定用于检查奇异性的容差值（使用 CRITERIA 子命令）。
- 为保存的变量创建用户指定名称（使用 SAVE 子命令）。
- 生成常规可估计函数表（使用 PRINT 子命令）。

请参阅命令语法参考以获取完整的语法信息。

---

## 第 11 章 复杂样本序数回归

“复杂样本序数回归”过程对通过复杂抽样方法抽取的样本的二元或有序因变量执行序数回归分析。您还可以请求对子群体进行分析。

**示例。** 议员在向立法院提交某项法案之前想了解公众是否支持该法案，以及对该法案的支持与选民人群统计信息有何关联。民意测验专家根据复杂抽样设计并实施了一些采访。使用复杂样本序数回归，可以根据选民人群统计信息将模型拟合度到对方案的支持水平。

复杂样本序数回归数据注意事项

**数据。** 因变量是序数变量。因子是分类变量。协变量是与因变量相关的定量变量。子群体变量可以是字符串或数值，但应该是分类变量。

**假设。** 数据文件中的个案代表来自复杂设计的一个样本，该样本应根据在“复杂样本计划”对话框中所选文件内的指定项进行分析。

获取复杂样本序数回归

1. 从菜单中选择:

**分析 > 复杂样本 > 序数回归...**

2. 选择计划文件。根据需要，选择自定义的联合概率文件。

3. 单击**继续**。

4. 在“复杂样本序数回归”对话框中，选择一个因变量。

根据需要，您可以:

- 为因子和协变量选择适合您数据的变量。
- 指定用于定义子群体的变量。尽管在整个数据集基础上对方差进行了正确估计，但仍只对所选的子群体变量类别进行分析。
- 选择关联函数。

**关联函数。** 关联函数是累积概率的转换形式，可用于模型估计。以下五个关联函数可用。

- **Logit**。  $f(x)=\log(x/(1-x))$ 。通常用于均匀分布的类别。
- **互补双对数**。  $f(x)=\log(-\log(1-x))$ 。通常在可能存在更多较高类别时使用。
- **负双对数**。  $f(x)=-\log(-\log(x))$ 。通常在可能存在更多较低类别时使用。
- **Probit**。  $f(x)=\Phi^{-1}(x)$ 。通常在潜变量正态分布时使用。
- **Cauchit (逆 Cauchy)**。  $f(x)=\tan(\pi(x-0.5))$ 。通常在潜变量有许多个极值时使用。

---

### 复杂样本序数回归: 响应概率

在“响应概率”对话框中，可以指定响应的累积概率是否随因变量值的增大或减小而增大，累积概率是属于因变量特定类别之前类别（包括该特定类别）的概率。

---

## 复杂样本序数回归：模型

**指定模型效应。** 缺省情况下，该过程使用主对话框中指定的因子和协变量构建主效应模型。此外，还可以构建包含交互效应和嵌套项的自定义模型。

非嵌套项

对于选定因子和协变量：

**交互。** 为所有选定变量创建最高级交互项。

**主效应。** 为每个选定的变量创建主效应项。

**所有二阶。** 创建选定变量的所有可能的双向交互。

**所有三阶。** 创建选定变量的所有可能的三阶交互。

**所有四阶。** 创建选定变量的所有可能的四阶交互。

**所有五阶。** 创建选定变量的所有可能的五阶交互。

嵌套项

在此过程中，可为您的模型建立嵌套项。嵌套项有助于对其值不与另一个因子的水平交互作用的因子或协变量的效应进行建模。例如，杂货连锁店可能在不同商店位置迎合顾客的不同消费习惯。由于每位顾客只经常光顾某一位置的商店，因此可以说客户效应**嵌套在**商店位置效应中。

此外，还可以包含交互效应，例如包含相同协变量的多项式项，或将多层嵌套添加到嵌套项。

**限制。** 嵌套项有以下限制：

- 一次交互内的所有因子必须是唯一的。因此，如果  $A$  是因子，则指定  $A*A$  是无效的。
- 嵌套效应内的所有因子必须是唯一的。因此，如果  $A$  是因子，则指定  $A(A)$  是无效的。
- 效应不可嵌套在协变量中。因此，如果  $A$  是因子且  $X$  是协变量，则指定  $A(X)$  是无效的。

---

## 复杂样本序数回归：统计

**模型拟合度。** 控制度量总体模型性能的统计的显示。

- **伪  $R^2$  方。** 在序数回归模型中，没有与线性回归的  $R^2$  对应的统计。相反，却有多种试图模拟  $R^2$  统计的属性的测量。
- **分类表。** 按因变量的模型预测类别显示观察类别的制表交叉分类。

**参数。** 使用此组可以控制与模型参数有关的统计的显示。

- **估算。** 显示系数的估计值。
- **取幂估值。** 显示以系数估值为幂的自然对数的底数。当该估值对于统计检验有良好的属性时，取幂估值（即  $\exp(B)$ ）更易于解释。
- **标准误差。** 显示每个系数估计值的标准误差。
- **置信区间。** 显示每个系数估计值的置信区间。在“选项”对话框中设置该区间的置信度。
- **T 检验。** 显示每个系数估计值的  $t$  检验。每个检验的原假设是该系数的值为 0。
- **参数估值协方差。** 显示模型系数的协方差矩阵的估计值。

- **参数估值的相关性。** 显示模型系数的相关性矩阵的估计值。
- **设计效果。** 估计值的方差与通过假设样本为简单随机样本所获得的方差的比率。这是指定复杂设计的效果测量，该值与 1 相差越大，表示效果越大。
- **设计效果的平方根。** 这是指定复杂设计的效果的测量值，以和标准误差单位相当的单位表示，值与 1 相差越大表示效果越好。

**平行线。** 在这一组中，可以请求与具有非平行线的模型关联的统计，其中分别为各个响应类别（最后一个除外）单独拟合一个回归线。

- **Wald 检验。** 生成回归参数对所有累积响应都相等的原假设检验。估计具有非平行线的模型并应用相等参数的 Wald 检验。
- **参数估计。** 显示具有非平行线的模型的系数和标准误差的估计值。
- **参数估值协方差。** 显示具有非平行线的模型的系数的协方差矩阵估计值。

**模型变量的汇总统计。** 显示有关因变量、协变量和因子的摘要信息。

**样本设计信息。** 显示有关样本的摘要信息，包括未加权的计数和群体大小。

## 复杂样本假设检验

**检验统计。** 在这一组中，可以选择用于检验假设的统计类型。可以在  $F$ 、调整的  $F$ 、卡方和调整的卡方之间选择。

**样本自由度。** 在这一组中，可以控制用于计算所有检验统计的  $p$  值的抽样设计自由度。如果基于抽样设计，该值为抽样第一阶段的主抽样单元格数和层数之差。或者，也可以通过指定一个正整数设置自定义自由度。

**调整的多重比较。** 在执行包含多重比较的假设检验时，总体显著性水平可从所包含的比較的显著性水平进行调节。使用此组可以选择调节方法。

- **显著性最低的差异。** 此方法并不控制拒绝某些线性对比不同于原假设值这一假设的总体概率。
- **连续 Sidak (Sequential Sidak).** 这是一个逐步下降的排斥性 Sidak 过程，就排斥单个假设而言，其保守性小得多，且保持了相同的总体显著性水平。
- **连续 Bonferroni (Sequential Bonferroni).** 这是按顺序逐步降低的拒绝 Bonferroni 过程，在拒绝个别假设方面不保守，但维持相同的总体显著性水平。
- **Sidak.** 此方法提供比 Bonferroni 方法更严格的界限。
- **Bonferroni.** 此方法根据多项对比检验调整观察到的显著性水平。

## 复杂样本序数回归：几率比

使用“几率比”对话框，可以控制指定的因子和协变量的模型估计累积几率比的显示。此功能仅用于使用 Logit 关联函数的模型。计算因变量所有类别（最后一个除外）的单个累积几率比；比例几率模型假设它们都相等。

**因子。** 对于所选的每个因子，都显示该因子处于每个类别的累积几率与处于指定参考类别的几率之比。

**协变量。** 对于所选的每个协变量，都显示协变量平均值加上指定变化单位的累积几率与处于平均值的几率之比。

在计算因子或协变量的几率比时，该过程将修正所有处于其最高级的其他因子以及所有处于其平均值的其他协变量。如果因子或协变量与模型中的其他预测变量交互作用，则几率比不仅取决于指定变量的变化，而且还取决于与之交互的变量的值。如果指定的协变量在模型中与其自身交互作用（例如  $age*age$ ），则几率比同时取决于协变量的变化和协变量的值。

---

## 复杂样本序数回归: 保存

**保存变量。** 在这一组中, 可以将模型预测类别、预测类别的概率、观察类别的概率、累积概率和预测概率保存为活动数据集中的新变量。

**将模型导出为 IBM SPSS Statistics 数据。** 写入一个 IBM SPSS Statistics 格式的数据集, 包含具有参数估计值、标准误差、显著性值和自由度的参数相关性或协方差矩阵。矩阵文件中变量顺序如下。

- **RowType\_**。取值 (或值标签) 为 COV (协方差)、CORR (相关性)、EST (参数估计)、SE (标准误差)、SIG (显著性水平) 和 DF (抽样设计自由度)。存在每个模型参数的 COV (或 CORR) 行类型的单独个案, 以及每个其他行类型的个案。
- **VarName\_**。对于行类型 COV 或 CORR, 取值为 P1、P2、..., 对应于所有模型参数的有序列表, 值标签对应于在参数估计值表中显示的参数字符串。对于其他行类型, 单元格为空。
- **P1, P2, ...** 这些变量对应于所有模型参数的有序列表, 值标签对应于在参数估计值表中显示的参数字符串, 这些变量根据行类型取值。对于冗余参数, 所有协方差设为零; 相关性设为系统缺失值; 所有参数估计值设为零; 并且所有标准误差、显著性水平和残差自由度设为系统缺失值。

注: 该文件不能立即用于在其他读取矩阵文件的过程中执行进一步分析, 除非这些过程接受在此导出的所有行类型。

**将模型导出为 XML。** 将参数估计值和参数协方差矩阵 (如果选择) 以 XML (PMML) 格式保存。您可以使用该模型文件以应用模型信息到其他数据文件用于评分目的。

---

## 复杂样本序数回归: 选项

**估计方法。** 可以选择参数估计方法; 其中包括: Newton-Raphson、Fisher 评分方法以及先执行 Fisher 评分迭代再切换为 Newton-Raphson 方法的混合方法。如果在混合方法的 Fisher 评分方法阶段期间, 在达到 Fisher 迭代的最大次数之前实现了收敛性, 则算法将继续执行 Newton-Raphson 方法。

**估值。** 在这一组中, 可以控制模型估计中使用的各种条件。

- **最大迭代次数。** 算法将执行的最大迭代次数。指定一个非负整数。
- **最大折半次数。** 每次迭代时, 步长都会减去因子 0.5, 直到对数似然估计增加或者达到最大步骤对分。指定一个正整数。
- **根据参数估值更改限制迭代。** 如果选择此项, 算法将在参数估计值的绝对或相对变化小于指定值 (必须为非负值) 的迭代之后停止。
- **根据对数似然估计更改限制迭代。** 如果选择此项, 算法将在对数似然估计函数的绝对或相对变化小于指定值 (必须为非负值) 的迭代之后停止。
- **检查数据点的完整分隔。** 如果选择此项, 算法将执行检验以确保参数估计值具有唯一值。当过程可生成一个正确对每个个案进行分类的模型时, 将发生分离。
- **显示迭代历史记录。** 从第 0 次迭代 (初始估计) 开始, 在每  $n$  次迭代时显示参数估计值和统计。如果选择打印迭代历史记录, 则无论  $n$  值为多少, 将总是打印最后一次迭代。

**用户缺失值。** 刻度设计变量以及因变量和任何协变量都应包含有效数据。从分析中删除含有任何这些变量的无效数据的个案。使用这些控制可以确定是否在层次、聚类、子群体和因子变量中将用户缺失值视为有效。

**置信区间。** 这是系数估值、取幂系数估值和几率比的置信区间度。指定大于等于 50 且小于 100 的值。

---

## CSORDINAL 命令附加功能

使用命令语法语言还可以:

- 指定作用对比作用的线性组合或一个值的自定义检验 (使用 CUSTOM 子命令)。
- 在计算因子和协变量的累积几率比 (使用 ODDSRATIOS 子命令) 时, 修正其他不处于其平均值的模型变量值。
- 在请求几率比 (使用 ODDSRATIOS 子命令) 时, 使用未标注值作为因子的自定义参考类别。
- 指定用于检查奇异性的容差值 (使用 CRITERIA 子命令)。
- 生成常规可估计函数表 (使用 PRINT 子命令)。
- 保存 25 个以上的概率变量 (使用 SAVE 子命令)。

请参阅命令语法参考以获取完整的语法信息。





## 第 12 章 复杂样本 Cox 回归

复杂样本 Cox 回归过程对由复杂抽样方法抽取的样本进行生存分析。您还可以请求对子群体进行分析。

**示例。** 政府执法机构关心其管辖区域内的屡犯率。测量屡犯率的方法之一就是罪犯第二次被捕的时间。机构希望利用 Cox 回归对时间建模以进行再次抓捕，但又担心比例风险假定在跨越年龄类别时失效。

医疗研究者正在对结束肌肉萎缩后中风症状复原计划后的患者存活时间进行调查。自从患者病史如所记录的显著非死亡事件的发生和次数而变化以来就存在每位主体具有多项个案的潜在性。就观察存活时间被复元长度“夸大”的意义而言，样本同样为左侧截断，因为当肌肉萎缩中风的风险攻击开始产生时，只有在复元计划后存活下来的患者存在于样本中。

### 复杂样本 Cox 回归数据注意事项

**生存时间。** 将 Cox 回归应用于存活时间分析的过程，即事件发生前的时间长度。有两种方式特定存活时间，取决于区间的开始时间：

- **时间=0。** 通常情况下，将在每个主体的区间开始具有完整信息并仅具有包含结束时间（或从“日期和时间”变量中创建一个带有结束时间的单一变量；请参阅以下内容）的变量。
- **依对象变化。** 这在您具有左侧截断，又称延迟条目时很恰当；例如，您正在对结束中风后症状复原计划后的患者存活时间进行分析，您可能会考虑到他们在中风时开始产生的风险攻击。但是，若您的样本只包含复元计划中存活的患者，就观察的存活时间被复元长度“夸大”的意义而言，则您的样本为左侧截断。可以通过将他们结束复元的时间指定为进入研究的时间对此进行说明。

**日期 & 时间变量。** “日期 & 时间”变量不能被用于直接界定区间的起始点与结束点；您若具有“日期 & 时间”就应该用其创建包含存活时间的变量。若无左侧截断，则根据进入研究的日期与观察日期期间的差异仅创建包含结束时间的变量。若无左侧截断，则根据开始研究的日期与进入日期期间的差异创建包含开始时间的变量并根据开始研究的日期与观察日期期间的差异创建包含结束时间的变量。

**事件状态。** 您需要记录了主体在区间中是否经历过被观察事件的变量。事件并非为其发生的主体为右侧已审查。

**对象标识。** 您可以通过拆分跨多项个案的单一主体观察轻松合并分段恒定、依时预测变量。例如，若您正在分析中风后患者存活时间，代表其医疗历史的变量应该是有用的预测变量。随着时间变化，他们可能会经历改变其医疗历史的重要医疗事件。下表显示了构造此类数据集的方法：*Patient ID* 为主体标识，*End time* 定义被观察的区间，*Status* 记录重要医疗事件，且 *Prior history of heart attack* 和 *Prior history of hemorrhaging* 为分段恒定、依时预测变量。

表 1. 用于合并分段恒定、依时预测变量的数据结构。

<i>Patient ID</i>	<i>End time</i>	状态	<i>Prior history of heart attack</i>	<i>Prior history of hemorrhaging</i>
1	5	心脏病	否(O)	否(O)
1	7	出血病	是(Y)	否(O)
1	8	死亡	是(Y)	是(Y)
2	24	死亡	否(O)	否(O)
3	8	心脏病	否(O)	否(O)
3	15	死亡	是(Y)	否(O)

**假设。** 数据文件中的个案代表来自复杂设计的一个样本，该样本应根据在“复杂样本计划”对话框中所选文件内的指定项进行分析。

通常情况下，Cox 回归模型假定比例风险，即从一个个案到另一个个案间的风险比不应随时间变化。若此假设不成立，可能需要将依时预测变量添加至模型。

**Kaplan-Meier 分析。** 您若不选择任何预测变量（或不向模型中输入任何已选预测变量）而选择在“选项”选项卡中利用产品限制方法计算基线存活曲线，则该过程将进行 Kaplan-Meier 型存活分析。

获取复杂样本 Cox 回归

1. 从菜单中选择：

分析 > 复杂样本 > Cox 回归...

2. 选择计划文件。根据需要，选择自定义的联合概率文件。
3. 单击**继续**。
4. 通过选择进入和退出研究的时间指定存活时间。
5. 选择事件状态变量。
6. 单击定义事件并定义至少一个事件值。

根据需要，可以选择主体标识。

---

## 定义事件

指定能指明终端事件已发生的值。

- **单一值。** 通过将其输入网格或从带有界定值标签的列表中将其选中以指定一个或多个值。
- **值范围。** 通过输入最小值和最大值或从带有界定值标签的列表中选择值以指定一系列值。

---

## 预测变量

“预测变量”选项卡可以指定用于建立模型效应的因子及协变量。

**因子。** 因子是分类预测变量，可以是数值或字符串。

**协变量。** 协变量为刻度预测变量，必须为数值。

**依时预测变量。** 存在某些比例风险假设不成立的情况。也就是说，风险比率随时间变化；在不同的时间点一个（或多个）预测变量的值会不同。在此情况下，需要指定依时预测变量。请参阅『界定依时预测变量』主题以获取更多信息。依时预测变量可作为因子或协变量选择。

## 界定依时预测变量

“界定依时预测变量”对话框可以创建依据内嵌时间变量的预测变量， $T_{i-}$ 。您可以使用此变量通过两种常用方法定义依时协变量：

- 若想对允许非成比例风险的扩展Cox 回归模型进行估计，可通过将依时预测变量界定为问题中的时间变量  $T_{i-}$  与协变量功能来实现。一个常见的例子就是简单地将时间变量和预测变量相乘，不过也可以指定较为复杂的函数。
- 有些变量在不同的时间段内可能具有不同的值，但其值与时间并不具有系统相关性。在这样的情况下，您需要定义一个**分段依时预测变量**，这可以通过使用逻辑表达式完成。逻辑表达式使用值 1 表示“true”，使用值 0 表示“false”。通过使用一系列逻辑表达式，您就可以使用一组度量创建依时预测变量。例如，如果您在

一个为期四周的研究中每周测量一次血压（使用 *BP1* 到 *BP4* 标识），则可以将依时预测变量定义为  $(T_ < 1) * BP1 + (T_ \geq 1 \ \& \ T_ < 2) * BP2 + (T_ \geq 2 \ \& \ T_ < 3) * BP3 + (T_ \geq 3 \ \& \ T_ < 4) * BP4$ 。注意，对于任何给定的个案，括号中都正好有一个项等于 1，其余项都等于 0。换言之，此函数意味着如果时间小于一周则使用 *BP1*；如果时间大于一周但小于两周则使用 *BP2*；依此类推。

注：若分段的依时预测变量如以上给出的血压示例一样与段恒定，则更容易通过拆分跨越多项个案的主体指定分段恒定、依时预测变量。有关更多信息，请参阅第 37 页的第 12 章，『复杂样本 Cox 回归』中针对“主体标识”的讨论。

在“界定依时预测变量”对话框中，您可以使用函数构建控件构建依时协变量的表达式，或者可以在“数字表达式”文本区域中直接输入表达式。注意，字符串常数必须包含在引号或单引号中，数字常数必须以美式格式键入，并使用句点作为小数定界符。结果变量被赋予指定的名称并应作为因子或协变量包含在“预测变量”选项卡中。

---

## 子组

**基线层。** 为每个变量值计算单独基线风险和生存功能的同时，跨层估计一个单一模型系数集。

**子群体变量。** 指定用于定义子群体的变量。仅对子群体变量的所选类别执行该分析。

---

## 模型

**指定模型效应。** 缺省情况下，该过程使用主对话框中指定的因子和协变量构建主效应模型。此外，还可以构建包含交互效应和嵌套项的自定义模型。

非嵌套项

对于选定因子和协变量：

**交互。** 为所有选定变量创建最高级交互项。

**主效应。** 为每个选定的变量创建主效应项。

**所有二阶。** 创建选定变量的所有可能的双向交互。

**所有三阶。** 创建选定变量的所有可能的三阶交互。

**所有四阶。** 创建选定变量的所有可能的四阶交互。

**所有五阶。** 创建选定变量的所有可能的五阶交互。

嵌套项

在此过程中，可为您的模型建立嵌套项。嵌套项有助于对其值不与另一个因子的水平交互作用的因子或协变量的效应进行建模。例如，杂货连锁店可能在不同商店位置迎合顾客的不同消费习惯。由于每位顾客只经常光顾某一位置的商店，因此可以说客户效应**嵌套在**商店位置效应中。

此外，还可以包含交互效应，例如包含相同协变量的多项式项，或将多层嵌套添加到嵌套项。

**限制。** 嵌套项有以下限制：

- 一次交互内的所有因子必须是唯一的。因此，如果 *A* 是因子，则指定 *A\*A* 是无效的。
- 嵌套效应内的所有因子必须是唯一的。因此，如果 *A* 是因子，则指定 *A(A)* 是无效的。

- 效应不可嵌套在协变量中。因此，如果  $A$  是因子且  $X$  是协变量，则指定  $A(X)$  是无效的。

---

## 统计

**样本设计信息。** 显示有关样本的摘要信息，包括未加权的计数和群体大小。

**事件与检查摘要。** 显示关于已审查个案数量与比例的摘要信息。

**事件时间中的风险设置。** 显示每个基线层中每个事件时间的事件数量和带风险数量。

**参数。** 使用此组可以控制与模型参数有关的统计的显示。

- **估算。** 显示系数的估计值。
- **取幂估值。** 显示以系数估值为幂的自然对数的底数。当该估值对于统计检验有良好的属性时，取幂估值（即  $\exp(B)$ ）更易于解释。
- **标准误差。** 显示每个系数估计值的标准误差。
- **置信区间。** 显示每个系数估计值的置信区间。在“选项”对话框中设置该区间的置信度。
- **t 检验。** 显示每个系数估计值的  $t$  检验。每个检验的原假设是该系数的值为 0。
- **参数估值协方差。** 显示模型系数的协方差矩阵的估计值。
- **参数估值的相关性。** 显示模型系数的相关性矩阵的估计值。
- **设计效果。** 估计值的方差与通过假设样本为简单随机样本所获得的方差的比率。这是指定复杂设计的效果测量，该值与 1 相差越大，表示效果越大。
- **设计效果的平方根。** 这是指定复杂设计的效果测量，该值与 1 相差越大，表示效果越大。

**模型假设。** 此组可生成对成比例风险假设的检测。该检测将拟合模型与包含依时预测变量  $x*_TF$  的备用模型做对比，每个预测变量  $x$  都具有  $_TF$  特定时间函数。

- **时间函数。** 指定备用模型  $_TF$  的形式。对于**恒等函数**， $_TF=T_$ 。对于**对数函数**， $_TF=T_$ 。对于**Kaplan-Meier**， $_TF=1-S_{KM}(T_)$ ，其中  $S_{KM}(\cdot)$  是生存函数的 Kaplan-Meier 估计。对于**等级**， $_TF$  在被观察的结束时间中是  $T_$  的等级顺序。
- **其他模型的参数估计值。** 在备用模型中显示每个参数的估计、标准误差和置信区间。
- **其他模型的协变量矩阵。** 在备用模型中显示参数间估计的协方差矩阵。

**基线生存和累积风险函数。** 连同其标准误差一同显示基线生存函数与基线累积风险。

注：若“预测变量”选项卡中界定的依时预测变量包含在模型中，则此选项不可用。

---

## 图

“图”选项卡可以申请风险函数图、生存函数图、对数负对数生存函数图以及 1 减生存函数图。还可以选择根据特定函数绘制置信区间图；在“选项”选项卡中设置置信度。

**预测变量模式。** 可以指定预测变量值模式用于要求图与“导出”选项卡中的导出生存文件。注意，若“预测变量”选项卡中界定的依时预测变量包含在模型中，则此选项不可用。

- **图表因子在。** 缺省情况下，每个因子以其最高水平进行评估。若有需要请输入或选择不同水平。或者，还可以选择通过为该因子选择复选框为单一因子的每个水平绘制分离线。
- **图表协变量在。** 每个协变量以其平均值进行评估。若有需要请输入或选择不同值。

---

## 假设检验

**检验统计。** 在这一组中，可以选择用于检验假设的统计类型。可以在  $F$ 、调整的  $F$ 、卡方和调整的卡方之间选择。

**样本自由度。** 在这一组中，可以控制用于计算所有检验统计的  $p$  值的抽样设计自由度。如果基于抽样设计，该值为抽样第一阶段的主抽样单元格数和层数之差。或者，也可以通过指定一个正整数设置自定义自由度。

**调整的多重比较。** 在执行包含多重比较的假设检验时，总体显著性水平可从所包含的比較的显著性水平进行调整。使用此组可以选择调节方法。

- **显著性最低的差异。** 此方法并不控制拒绝某些线性对比不同于原假设值这一假设的总体概率。
- **连续 Sidak (Sequential Sidak).** 这是一个逐步下降的排斥性 Sidak 过程，就排斥单个假设而言，其保守性小得多，且保持了相同的总体显著性水平。
- **连续 Bonferroni (Sequential Bonferroni).** 这是按顺序逐步降低的拒绝 Bonferroni 过程，在拒绝个别假设方面不保守，但维持相同的总体显著性水平。
- **Sidak.** 此方法提供比 Bonferroni 方法更严格的界限。
- **Bonferroni.** 此方法根据多项对比检验调整观察到的显著性水平。

---

## 保存

**保存变量。** 此组可以将相关模型变量保存至活动数据集以备将来用于诊断和结果报告。注意，当依时预测变量包含在模型中时，全部不可用。

- **生存函数。** 以被观察时间和每个个案的预测变量值保存生存概率（生存函数值）。
- **生存函数的置信区间下限。** 以被观察时间和每个个案的预测变量值保存生存函数的置信区间下限。
- **生存函数的置信区间上限。** 以被观察时间和每个个案的预测变量值保存生存函数的置信区间上限。
- **累积风险函数。** 以被观察时间和每个个案的预测变量值保存累积风险或  $\ln(\text{survival})$ 。
- **累积风险函数的置信区间下限。** 以被观察时间和每个个案的预测变量值保存累积风险函数的置信区间下限。
- **累积风险函数的置信区间上限。** 以被观察时间和每个个案的预测变量值保存累积风险函数的置信区间上限。
- **线性预测变量的预测值。** 保存参考值的线性组合，更正预测变量时间回归系数。线性预测变量是风险函数占基线风险的比率。在成比例风险模型下，此值为时间恒定值。
- **Schoenfeld 残差。** 对于模型中的每个为审查个案和非冗余参数，Schoenfeld 残差是与模型参数相关的被观察预测变量值与以被观察时间设置的带风险个案预期预测变量值之间的差。Schoenfeld 残差可用于帮助使用成比例风险假设；例如，对于预测变量  $x$ ，若成比例风险成立，依时预测变量  $x \cdot \ln(T_0)$  与时间的 Schoenfeld 残差图应该在 0 处显示一条水平线。为模型中的每个非冗余参数保存单独变量。Schoenfeld 残差只为为审查个案计算。
- **Martingale 残差。** 对于每个个案，martingale 残差是被观察审查（审查为 0，未审查为 1）与观察时间内的的事件预期值之间的差。
- **偏差残差。** 偏差残差是 martingale 残差“调整”后关于 0 更加对称的表现。偏差残差对于预测变量的图不应揭示任何模式。
- **Cox-Snell 残差。** 对于每个个案，Cox-Snell 残差是观察时间内的的事件预期值或被观察审查减去 martingale 残差。
- **Score 残差。** 对于每个个案和模型中的冗余参数，score 残差是第一个伪似然导数个案的贡献。为模型中的每个非冗余参数保存单独变量。

- **DFBeta 残差。** 对于每个个案和模型中的冗余参数，当个案从模型中被删除时，DFBeta 残差近似于参数估计值的变化。带有大型 DFBeta 残差的个案可能会对分析施加过度影响。为模型中的每个非冗余参数保存单独变量。
- **汇总残差。** 当多项个案代表一个单一主体时，主体的汇总残差只是全部个案中属于同一主体的相应个案残差的总和。对于 Schoenfeld 残差，汇总版本与非汇总版本相同，因为 Schoenfeld 残差只为未审查个案定义。这些残差只有当主体标识在“时间与事件”选项卡中被指定时方可使用。

**保存的变量名称。** 自动名称生成以确保您能保存您的所有工作。无需先删除数据编辑器中保存的变量，自定义名称允许您放弃/替换上一次运行的结果。

---

## 导出

**将模型导出为 IBM SPSS Statistics 数据。** 写入一个 IBM SPSS Statistics 格式的数据集，包含具有参数估计值、标准误差、显著性值和自由度的参数相关性或协方差矩阵。矩阵文件中变量顺序如下。

- **RowType\_。** 取值（或值标签）为 COV（协方差）、CORR（相关性）、EST（参数估计）、SE（标准误差）、SIG（显著性水平）和 DF（抽样设计自由度）。存在每个模型参数的 COV（或 CORR）行类型的单独个案，以及每个其他行类型的个案。
- **VarName\_。** 对于行类型 COV 或 CORR，取值为 P1、P2、...，对应于所有模型参数的有序列表，值标签对应于在参数估计值表中显示的参数字符串。对于其他行类型，单元格为空。
- **P1, P2, ...** 这些变量对应于所有模型参数的有序列表，值标签对应于在参数估计值表中显示的参数字符串，这些变量根据行类型取值。对于冗余参数，所有协方差设为零；相关性设为系统缺失值；所有参数估计值设为零；并且所有标准误差、显著性水平和残差自由度设为系统缺失值。

注：该文件不能立即用于在其他读取矩阵文件的过程中执行进一步分析，除非这些过程接受在此导出的所有行类型。

**将生存函数导出为 IBM SPSS Statistics 数据。** 为每个失败或事件时间编写一个包含生存函数的 IBM SPSS Statistics 格式数据集；生存函数标准误差；生存函数置信区间上下限；以及累积风险函数，并以基线和在“图”选项卡中指定的预测变量模式对其进行评估。矩阵文件中变量顺序如下。

- **基线层变量。** 为每个分层变量值生成单独的生存分析表。
- **生存时间变量。** 事件时间；为每个独有的事件时间创建一个单独个案。
- **Sur\_0, LCL\_Sur\_0, UCL\_Sur\_0.** 基线生存函数与其置信区间的上下限。
- **Sur\_R, LCL\_Sur\_R, UCL\_Sur\_R.** 以“参考”模式评估的生存函数（请参阅输出中的模式值表格）与其置信区间的上下限。
- **Sur\_#.#, LCL\_Sur\_#.#, UCL\_Sur\_#.#, ...** 以在“图”选项卡中指定的每种预测变量模式评估的生存函数与其置信区间的上下限。请参阅输出中的模式值表格，与带有数字 #.# 的模式匹配。
- **Haz\_0, LCL\_Haz\_0, UCL\_Haz\_0.** 基线累积风险函数与其置信区间的上下限。
- **Haz\_R, LCL\_Haz\_R, UCL\_Haz\_R.** 以“参考”模式评估的累积风险函数（请参阅输出中的模式值表格）与其置信区间的上下限。
- **Haz\_#.#, LCL\_Haz\_#.#, UCL\_Haz\_#.#, ...** 以在“图”选项卡中指定的每种预测变量模式评估的累积风险函数与其置信区间的上下限。请参阅输出中的模式值表格，与带有数字 #.# 的模式匹配。

**将模型导出为 XML。** 保存预测生存函数所需的全部信息，包括 XML（PMML）格式的参数评估与基线生存函数。您可以使用该模型文件以应用模型信息到其他数据文件用于评分目的。

---

## 选项

**估值。** 这些控件为回归系数的评估指定准则。

- **最大迭代次数。** 算法将执行的最大迭代次数。指定一个非负整数。
- **最大折半次数。** 每次迭代时，步长都会减去因子 0.5，直到对数似然估计增加或者达到最大步骤对分。指定一个正整数。
- **根据参数估值更改限制迭代。** 如果选择此项，算法将在参数估计值的绝对或相对更改小于指定值（必须为正值）的迭代之后停止。
- **根据对数似然估计更改限制迭代。** 如果选择此项，算法将在对数似然估计函数的绝对或相对更改小于指定值（必须为正值）的迭代之后停止。
- **显示迭代历史记录。** 显示参数估计值和伪对数似然估计的迭代历史记录，并打印对参数评估变化和伪对数似然的最后一次评估。迭代历史记录表打印从第 0 次迭代（初始估计值）开始的每  $n$  次迭代，其中  $n$  代表增量值。如果请求迭代历史记录，那么无论  $n$  的值是多少，都会显示最后一次迭代。
- **参数估计的断开连接方法。** 当出现绑定的被观察失败时间时，此种方法之一将用于断开绑定。Efron 方法为更加大量的计算。

**生存函数。** 这些控件为涉及生存函数的计算指定准则。

- **基线生存函数的估算方法。** **Breslow**（或 Nelson-Aalan 或经验的）方法通过非减少步骤函数与被观察失败时间共同评估基线累积风险，然后通过关系  $\text{survival}=\exp(\text{cumulative hazard})$  计算基线生存。**Efron** 方法为更加大量的计算并当无绑定时简化为 Breslow 方法。**产品限制**方法通过非增加的正确连续函数评估基线生存；当模型中不存在任何预测变量时，此方法简化为 Kaplan-Meier 评估。
- **生存函数的置信区间。** 可通过以下三种方式对置信区间进行计算：在初始单位中、通过对数转换，或对数负对数转换。只有对数负对数转换能保证置信区间介于 0 和 1 之间，但对数转换被普遍认为是性能“最好”的方法。

**用户缺失值。** 要在分析中包含个案，所有变量必须具有有效值。这些控件可以决定用户缺失值在类别模型（包括因子、事件、层次和子群体变量）和取样设计变量中是否有效。

**置信区间 (%)。** 这是用于系数估计值、取幂系数估计值以及累积风险函数估计值的置信区间水平。指定大于等于 0 且小于 100 的值。

---

## CSCOXREG 命令的附加功能

使用命令语言还可以：

- 执行自定义的假设检验（使用 CUSTOM 子命令和 /PRINT LMATRIX）。
- 容差规范（使用 /CRITERIA SINGULAR）。
- 一般可估计函数表格（使用 /PRINT GEF）。
- 多种预测变量模式（使用多个 PATTERN 子集）。
- 当根名称被指定时（使用 SAVE 子命令），将已保存变量的数量最大化。对话框接受 CSCOXREG 25 个变量的缺省。

请参阅命令语法参考以获取完整的语法信息。





---

## 声明

本信息是为在美国提供的产品和服务编写的。

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。有关您当前所在区域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。不过，用户应自行负责评估和验证任何非 IBM 产品、程序或服务的工作情况。

IBM 公司可能已拥有或正在申请与本文档内容有关的各项专利。您获得该文档并不意味着授予您任何这些专利许可。您可以将书面许可查询函件发送至：

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

本条款不适用英国或任何这样的条款与当地法律不一致的国家或地区：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。有些州/省不允许针对某些交易的明示或暗示免责条款，因此本声明可能不适用于您。

本信息可能含有技术误差或排版错误。此处的信息会定期进行更改；这些更改会体现在本出版物的新版本中。IBM 可以随时对本资料中描述的产品和/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）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 Software Group  
ATTN: Licensing

200 W. Madison St.  
Chicago, IL; 60606  
U.S.A.

此类信息的提供应遵照相关条款和条件，其中包括在某些情况下支付适当费用。

本资料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 国际软件许可协议或任何同等协议中的条款提供。

此处所含的性能数据均在受控环境下决定。因此，在其他操作环境中获得的结果可能差异较大。有些测量可能在开发级的系统中进行，不保证这些测量结果与常用系统上的测量结果相同。此外，有些测量结果可能通过推断来估计得出。实际结果可能有所差异。此文档的用户应针对其具体环境验证适用的数据。

非 IBM 产品的相关信息来自这些产品的供应商，及其发布的公告或其他公开来源。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。

有关 IBM 未来方向或意向的所有声明均可能未经通知即变更或撤销，并且仅代表目标和目的。

本信息包含日常业务运营中使用的数据和报告的示例。为了尽可能详尽地对其进行说明，示例中包含了人员的姓名、公司、品牌和产品的名称。所有这些名称均为虚构，与真实商业企业使用的名称和地址的任何雷同纯属巧合。

版权许可：

本信息包括源语言形式的样本应用程序，这些样本说明不同操作平台上的编程方法。如果是为按照在编写样本程序的操作平台上的应用程序编程接口 (API) 进行应用程序的开发、使用、经销或分发为目的，您可以任何形式对这些样本程序进行复制、修改、分发，而无须向 IBM 付费。这些示例并未在所有条件下作全面测试。因此，IBM 不能担保或暗示这些程序的可靠性、可维护性或功能。本样本程序仍然是“按现状”提供的，不附有任何种类的保证。对于因使用样本程序所引起的任何损害，IBM 概不负责。

凡这些实例程序的每份拷贝或其任何部分或任何衍生产品，都必须包括如下版权声明：

©（贵公司的名称）（年）。此部分代码是根据 IBM Corp. 公司的样本程序衍生出来的。

© Copyright IBM Corp. \_（输入年份）\_ . All rights reserved.

---

## 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。当前的 IBM 商标列表，可从 Web 站点 [www.ibm.com/legal/copytrade.shtml](http://www.ibm.com/legal/copytrade.shtml) 上“版权和商标信息”部分获取。

Adobe、Adobe 徽标、PostScript 和 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 或其子公司在美国和其他国家或地区的商标或注册商标。

Linux 是 Linus Torvalds 在美国、其他国家或地区或两者的注册商标。

Microsoft、Windows、Windows NT 和 Windows 徽标是 Microsoft Corporation 在美国、其他国家或地区或两者的商标。

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 及/或其分支机构的商标和注册商标。



# 索引

## [ B ]

包含概率  
    在抽样向导中 5  
比例风险测试  
    在“复杂样本 Cox 回归”中 40  
变异系数 (COV)  
    累计值 15  
    缺失值 17  
    危险度差值 19  
    在“复杂样本比率”中 21  
标准误差  
    累计值 15  
    缺失值 17  
    危险度差值 19  
    在“复杂样本比率”中 21  
    在“复杂样本序数回归”中 32  
    在“复杂样本一般线性模型”中 24  
    在“复杂样本 Logistic 回归”中 28  
步骤对分  
    在“复杂样本序数回归”中 34  
    在“复杂样本 Logistic 回归”中 30

## [ C ]

采样  
    复杂设计 3  
参考类别  
    在“复杂样本一般线性模型”中 25  
    在“复杂样本 Logistic 回归”中 27  
参数估计值  
    在“复杂样本序数回归”中 32  
    在“复杂样本一般线性模型”中 24  
    在“复杂样本 Cox 回归”中 40  
    在“复杂样本 Logistic 回归”中 28  
参数估值相关性  
    在“复杂样本序数回归”中 32  
    在“复杂样本一般线性模型”中 24  
    在“复杂样本 Logistic 回归”中 28  
参数估值协方差  
    在“复杂样本序数回归”中 32  
    在“复杂样本一般线性模型”中 24  
    在“复杂样本 Logistic 回归”中 28  
参数收敛  
    在“复杂样本序数回归”中 34  
    在“复杂样本 Logistic 回归”中 30  
残差  
    危险度差值 19  
    在“复杂样本一般线性模型”中 25  
差分对比  
    在“复杂样本一般线性模型”中 25

重复对比  
    在“复杂样本一般线性模型”中 25  
抽样方法  
    在抽样向导中 4  
抽样估计  
    在“分析准备向导”中 10

## [ D ]

大小测量  
    在抽样向导中 4  
调整的卡方  
    在“复杂样本 Cox 回归”中 41  
    在“复杂样本”中 24, 29, 33  
调整的 F 统计  
    在“复杂样本 Cox 回归”中 41  
    在“复杂样本”中 24, 29, 33  
迭代  
    在“复杂样本序数回归”中 34  
    在“复杂样本 Logistic 回归”中 30  
迭代历史记录  
    在“复杂样本序数回归”中 34  
    在“复杂样本 Logistic 回归”中 30  
对比  
    在“复杂样本一般线性模型”中 25  
多项式对比  
    在“复杂样本一般线性模型”中 25

## [ F ]

分层  
    在抽样向导中 3  
    在“分析准备向导”中 9  
分类表  
    在“复杂样本序数回归”中 32  
    在“复杂样本 Logistic 回归”中 28  
分离  
    在“复杂样本序数回归”中 34  
    在“复杂样本 Logistic 回归”中 30  
分析计划 9  
复杂抽样  
    分析计划 9  
    假设检验 24, 29, 33  
    缺失值 16, 20  
    选项 16, 18, 20, 22  
    样本计划 3  
复杂样本比率 21  
    缺失值 21  
    统计 21  
复杂样本交叉表 19  
    统计 19

复杂样本描述 17  
    缺失值 18  
    统计 17  
复杂样本频率 15  
    统计 15  
复杂样本序数回归 31  
    保存变量 34  
    模型 32  
    统计 32  
    响应概率 31  
    选项 34  
    “复杂样本交叉表格”中的 33  
复杂样本一般线性模型 23  
    保存变量 25  
    估计平均值 25  
    命令附加功能 26  
    模型 23  
    统计 24  
    选项 26  
复杂样本 Cox 回归  
    保存变量 41  
    定义事件 38  
    假设检验 41  
    模型 39  
    模型导出 42  
    日期和时间变量 37  
    统计 40  
    图 40  
    选项 43  
    依时预测变量 38  
    预测变量 38  
    子组 39  
    Kaplan-Meier 分析 37  
复杂样本 Logistic 回归 27  
    保存变量 29  
    参考类别 27  
    命令附加功能 30  
    模型 27  
    统计 28  
    选项 30  
    “复杂样本交叉表格”中的 29

## [ G ]

估计边际平均值  
    在“复杂样本一般线性模型”中 25

## [ H ]

行百分比  
    危险度差值 19

合计  
    缺失值 17  
汇总残差  
    在“复杂样本 Cox 回归”中 41

## [ J ]

基线层  
    在“复杂样本 Cox 回归”中 39  
计划文件 2  
简单对比  
    在“复杂样本一般线性模型”中 25  
简单随机抽样  
    在抽样向导中 4  
聚类  
    在抽样向导中 3  
    在“分析准备向导”中 9

## [ K ]

卡方统计  
    在“复杂样本 Cox 回归”中 41  
    在“复杂样本”中 24, 29, 33

## [ L ]

累积概率  
    在“复杂样本序数回归”中 34  
连续 Sidak 修正  
    在“复杂样本 Cox 回归”中 41  
    在“复杂样本”中 24, 29, 33  
列百分比  
    危险度差值 19

## [ P ]

偏差残差  
    在“复杂样本 Cox 回归”中 41  
偏移对比  
    在“复杂样本一般线性模型”中 25  
平行线检验  
    在“复杂样本序数回归”中 32  
平均值  
    缺失值 17

## [ Q ]

期望值  
    危险度差值 19  
缺失值  
    缺失值 18  
    在“复杂样本比率”中 21  
    在“复杂样本序数回归”中 34  
    在“复杂样本一般线性模型”中 26

缺失值 (续)  
    在“复杂样本 Logistic 回归”中 30  
    在“复杂样本”中 16, 20

## [ S ]

设计效果  
    累计值 15  
    缺失值 17  
    危险度差值 19  
    在“复杂样本比率”中 21  
    在“复杂样本序数回归”中 32  
    在“复杂样本一般线性模型”中 24  
    在“复杂样本 Cox 回归”中 40  
    在“复杂样本 Logistic 回归”中 28  
设计效果的平方根  
    累计值 15  
    缺失值 17  
    危险度差值 19  
    在“复杂样本比率”中 21  
    在“复杂样本序数回归”中 32  
    在“复杂样本一般线性模型”中 24  
    在“复杂样本 Cox 回归”中 40  
    在“复杂样本 Logistic 回归”中 28  
输入样本权重  
    在抽样向导中 3  
顺序抽样  
    在抽样向导中 4  
顺序 Bonferroni 修正  
    在“复杂样本 Cox 回归”中 41  
    在“复杂样本”中 24, 29, 33  
似然估计收敛性  
    在“复杂样本序数回归”中 34  
    在“复杂样本 Logistic 回归”中 30

## [ W ]

伪 R2 统计  
    在“复杂样本序数回归”中 32  
    在“复杂样本 Logistic 回归”中 28

## [ X ]

系统抽样  
    在抽样向导中 4  
相对风险  
    危险度差值 19  
响应概率  
    在“复杂样本序数回归”中 31

## [ Y ]

样本比例  
    在抽样向导中 5

样本大小  
    在抽样向导中 4, 5  
样本计划 3  
样本权重  
    在抽样向导中 5  
    在“分析准备向导”中 9  
样本设计信息  
    在“复杂样本 Cox 回归”中 40  
依时预测变量  
    在“复杂样本 Cox 回归”中 38  
预测概率  
    在“复杂样本序数回归”中 34  
    在“复杂样本 Logistic 回归”中 29  
预测类别  
    在“复杂样本序数回归”中 34  
    在“复杂样本 Logistic 回归”中 29  
预测值  
    在“复杂样本一般线性模型”中 25

## [ Z ]

置信度  
    在“复杂样本序数回归”中 34  
    在“复杂样本 Logistic 回归”中 30  
置信区间  
    累计值 15  
    缺失值 17  
    危险度差值 19  
    在“复杂样本比率”中 21  
    在“复杂样本序数回归”中 32  
    在“复杂样本一般线性模型”中 24, 26  
    在“复杂样本 Logistic 回归”中 28  
子群体  
    在“复杂样本 Cox 回归”中 39  
自由度  
    在“复杂样本 Cox 回归”中 41  
    在“复杂样本”中 24, 29, 33  
最小显著性差异  
    在“复杂样本 Cox 回归”中 41  
    在“复杂样本”中 24, 29, 33

## B

Bonferroni  
    在“复杂样本 Cox 回归”中 41  
    在“复杂样本”中 24, 29, 33  
Breslow 估计方法  
    在“复杂样本 Cox 回归”中 43  
Brewer 抽样方法  
    在抽样向导中 4

## C

Cox-Snell 残差  
    在“复杂样本 Cox 回归”中 41

## E

- Efron 估计方法  
在“复杂样本 Cox 回归”中 43

## F

- F 统计  
在“复杂样本 Cox 回归”中 41  
在“复杂样本”中 24, 29, 33
- Fisher 评分方法  
在“复杂样本序数回归”中 34

## H

- Helmert 对比  
在“复杂样本一般线性模型”中 25

## M

- martingale 残差  
在“复杂样本 Cox 回归”中 41
- Murthy 抽样方法  
在抽样向导中 4

## N

- Newton-Raphson 方法  
在“复杂样本序数回归”中 34

## P

- PPS 抽样  
在抽样向导中 4

## R

- R<sup>2</sup> 统计  
在“复杂样本一般线性模型”中 24

## S

- Sampford 抽样方法  
在抽样向导中 4
- Schoenfeld 偏残差  
在“复杂样本 Cox 回归”中 41
- score 残差  
在“复杂样本 Cox 回归”中 41
- Sidak 修正  
在“复杂样本 Cox 回归”中 41  
在“复杂样本”中 24, 29, 33

## T

- t 检验  
在“复杂样本序数回归”中 32  
在“复杂样本一般线性模型”中 24  
在“复杂样本 Logistic 回归”中 28

## [ 特别字符 ]

- “复杂样本交叉表格”中的  
危险度差值 19  
在“复杂样本序数回归”中 33  
在“复杂样本 Logistic 回归”中 29
- “复杂样本频率”中的  
累计值 15  
缺失值 17  
危险度差值 19  
在抽样向导中 5  
在“复杂样本比率”中 21









Printed in China