IBM SPSS Advanced Statistics 19

Note: Before using this information and the product it supports, read the general information under Notices第 143 页码.

前言

IBM® SPSS® Statistics 是一种用于分析数据的综合系统。Advanced Statistics 可选附加模块提供本手册中描述的其他分析方法。此 Advanced Statistics 附加模块必须与 SPSS Statistics Core 系统一起使用，并已完全集成到了该系统中。

关于 SPSS Inc.，IBM 下属公司

SPSS Inc. 是一家 IBM 下属公司，它也是全球领先的预测分析软件和解决方案提供商。该公司拥有全面的产品系列，涵盖数据收集、统计量、建模和部署，通过在业务流程中嵌入分析技术，收集人们的态度与看法，预测未来客户交互结果，然后针对这些深入见解采取相应行动。SPSS Inc. 解决方案着眼于整合分析技术、IT 基础设施和业务流程，以帮助达成整个企业内相互关联的业务目标。全球各地的众多企业、政府和学术机构客户依靠 SPSS Inc. 技术在吸引、留住和发展客户方面取得竞争优势，同时减少欺诈并缓解风险。SPSS Inc. 在 2009 年 10 月被 IBM 并购。有关更多信息，请访问 http://www.spss.com。

技术支持

我们提供有“技术支持”以维护客户。客户可就 SPSS Inc. 产品使用或某一受支持硬件环境的安装帮助寻求技术支持。要获得“技术支持”，请访问 SPSS Inc. 网站 http://support.spss.com，或通过网站 http://support.spss.com/default.asp?refpage=contactus.asp 找到当地办事处。在请求协助时，请准备好您和您组织的 ID 以及支持协议。

客户服务

如果对发货或帐户存在任何问题，请联系您当地的办事处，联系方式列在 Web 站点中，网址为 http://www.spss.com/worldwide。请先准备好您的序列号以供识别。

培训讲座

SPSS Inc. 提供公开的以及现场的培训讲座。所有讲座都是以实践小组为特色的。讲座将定期在各大城市开展。关于这些讲座的更多信息，请联系您本地的办事处，联系方式列在 Web 站点上，网址为 http://www.spss.com/worldwide。

附加出版物

SPSS Statistics：数据分析指南、SPSS Statistics：Statistical Procedures Companion 和 SPSS Statistics：Advanced Statistical Procedures Companion（由 Marija Norušis 编写，并已由 Prentice Hall 出版）作为建议的补充材料提供。这些出版物涵盖 SPSS Statistics Base 模块、Advanced Statistics 模块和 回归模块中的统计过程。无论您是刚开始从事数据分析工作，还是已准备好使用高级应用程序，这些书籍都将帮助您最有效地利用在 IBM® SPSS® Statistics 产品中找到的功能。有关其他信息，包括出版物的内容和示例章节，请参阅作者的网站： http://www.norusis.com

内容

章 1

Advanced Statistics 简介

Advanced Statistics 选项提供的过程可以提供比 Statistics Base 选项更高级的建模选项。

- “GLM 多变量”对“GLM 单变量”提供的一般线性模型进行了扩展，以允许使用多个因变量。更进一步的扩展“GLM 重复测量”允许重复度量多个因变量。
- “方差成分分析”是将因变量的变异性分解为固定和随机成分的特定工具。
- “线性混合模型”对一般线性模型进行了扩展，因此允许数据表现出相关的和不恒定的变异性。因此，线性混合模型提供了不仅能够就数据的均值还能够就其方差和协方差建模的灵活性。
- “广义线性模型”(GZLM) 放宽了误差项的正态假设，仅要求因变量通过转换或关联函数与预测变量线性相关。“广义估计方程”(GEE) 对 GZLM 进行了扩展，以允许重复度量。
- “一般对数线性分析”允许您为交叉分类计数数据拟合模型，“模型选择对数线性分析”可帮助您在模型间选择。
- “Logit 对数线性分析”允许您拟合对数线性模型来分析分类因变量与一个或多个分类预测变量之间的关系。
- “生存”分析通过“寿命表”提供，用于检查时间事件变量的分布，可能按因子变量水平分析；“Kaplan-Meier”生存分析用于检查时间事件变量的分布，可能按因子变量水平分析，或按分层变量水平产生单独的分析；“Cox 回归”用于根据给定协变量的值对指定事件的时间建模。

章
2

GLM 多变量分析

“GLM 多变量”过程通过一个或多个因子变量或协变量为多个因变量提供回归分析和方差分析。因子变量将总体划分成组。通过使用此一般线性模型过程，您可以检验关于因子变量对因变量联合分布的各个分组的均值的效应的原假设。可以调查因子之间的交互以及单个因子的效应。另外，还可以包含协变量的效应以及协变量与因子的交互。对于回归分析，自变量（预测变量）指定为协变量。

平衡与非平衡模型均可进行检验。如果模型中的每个单元包含相同的个案数，则设计是平衡的。在多变量模型中，模型中的效应引起的平方和以及误差平方和以矩阵形式表示，而不是以单变量分析中的标量形式表示。这些矩阵称为 SSCP（平方和与叉积）矩阵。如果指定了多个因变量，则提供使用 Pillai 的轨迹、Wilks 的 lambda、Hotelling 的轨迹、Roy 的最大根条件以及近似 F 统计量的多变量方差分析，同时还提供每个因变量的单变量方差分析。除了检验假设，“GLM 多变量”过程还生成参数估计。

常用的先验对比可用于执行假设检验。另外，在整体的 F 检验已显示显著性之后，可以使用两两比较检验评估指定均值之间的差值。估计边际均值为模型中的单元提供了预测均值估计值，且这些均值的轮廓图（交互图）允许您轻松对其中一些关系进行可视化。单独为每个因变量执行两两多重比较检验。

残差、预测值、Cook 距离以及杠杆值可以另存为数据文件中检查假设的新变量。另外还提供残差 SSCP 矩阵（残差的平方和与叉积的方形矩阵）、残差协方差矩阵（残差 SSCP 矩阵除以残差的自由度）和残差相关矩阵（残差协方差矩阵的标准化形式）。

WLS 权重允许您指定一个变量，用来针对加权最小平方 (WLS) 分析为观察值赋予不同权重，这样也许可以补偿测量的不同精确度。

示例。 某塑料制造商要测量塑料膜的三种属性：抗扯强度、光泽度和不透明度。厂商使用两种挤出速度和添加剂量进行了尝试，并对挤出速度和添加剂量的各种组合度量了这三种属性。厂商发现挤出速度和添加剂量单独产生的结果很明显，但这两种因子的交互作用并不明显。

方法。类型 I、类型 II、类型 III 和类型 IV 的平方和可用来评估不同的假设。类型 III 是缺省值。

统计量。两两比较范围检验和多重比较：最小显著性差异、Bonferroni、Sidak、Scheffé、Ryan-Einot-Gabriel-Welsch 多重 F、Ryan-Einot-Gabriel-Welsch 多范围、Student-Newman-Keuls、Tukey’s 真实显著性差异、Tukey 的 b、Duncan、Hochberg’s GT2、Gabriel、Waller-Duncan t 检验、Dunnett（单侧和双侧）、Tamhane’s T2、Dunnett’s T3、Games-Howell 和 Dunnett’s C。描述统计：所有单元中所有因变量的观察均值、标准差和计数；Levene 的方差齐性检验；对因变量协方差矩阵的齐性 Box 的 M 检验以及 Bartlett 的球形度检验。

图。分布-水平图、残差图以及轮廓图（交互）。

数据。因变量应是定量的。因子应是分类因子，可以具有数字值或字符串值。协变量是与因变量相关的定量变量。

假设。对于因变量，数据是来自多变量正态总体的随机向量样本；在总体中，所有单元的方差-协方差矩阵均相同。尽管数据应对称，但方差分析对于偏离正态性是稳健的。要检查假设，您可以使用方差齐性检验（包括 Box 的 M 检验）和分布-水平图。您还可以检查残差和残差图。

相关过程。在进行方差分析之前使用“探索”过程来检查数据。对于单个因变量，请使用“GLM 单变量”。如果您针对每个主体的多种情况度量相同的因变量，请使用“GLM 重复测量”。

获取 GLM 多变量表

▶ 从菜单中选择：

分析 > 一般线性模型 > 多变量...

图片 2-1
“多变量”对话框

▶ 请选择至少两个因变量。

或者，您也可以指定“固定因子”、“协变量”和“WLS 权重”。

GLM 多变量：模型

图片 2-2
“多变量：模型”对话框

指定模型。全因子模型包含所有因子主效应、所有协变量主效应以及所有因子间交互。它不包含协变量交互。选择**定制**可以仅指定其中一部分的交互或指定因子协变量交互。必须指定要包含在模型中的所有项。

因子与协变量。列出因子与协变量。

模型。模型取决于数据的性质。选择**定制**之后，您可以选择分析中感兴趣的主效应和交互效应。

平方和。计算平方和的方法。对于没有缺失单元的平衡或非平衡模型，类型 III 平方和方法最常用。

在模型中包含截距。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。

构建项

对于选定因子和协变量：

交互。创建所有选定变量的最高级交互项。这是缺省值。

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二阶交互。

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。

平方和

对于该模型，您可以选择平方和类型。类型 III 最常用，并且是缺省类型。

类型 I。此方法也称为平方和分级解构法。在模型中，每一项只针对它前面的那项进行调整。类型 I 平方和常用于：

- 平衡 ANOVA 模型，其中任何主效应在任何一阶交互效应之前指定，任何一阶交互效应在任何二阶交互效应之前指定，依此类推。
- 多项式回归模型，其中任何低阶项在任何高阶项之前指定。
- 纯嵌套模型，其中第一个指定的效应嵌套在第二个指定的效应中，第二个指定的效应嵌套在第三个指定的效应中，依此类推。（此嵌套形式只能通过使用语法来指定。）

类型 II。此方法在为所有其它“相应的”效应进行调节的模型中计算某个效应的平方和。相应的效应是指，与所有效应（不包含正被检查的效应）相对应的效应。类型 II 平方和方法常用于：

- 平衡 ANOVA 模型。
- 任何只有主要因子效应的模型。
- 任何回归模型。
- 纯嵌套设计。（此嵌套形式能通过使用语法来指定。）

类型 III。缺省类型。此方法在设计中通过以下形式计算某个效应的平方和：为任何不包含该效应的其他效应，以及任何与包含该效应正交的效应（如果存在）调整的平方和。类型 III 平方和具有一个主要优点，那就是只要可估计性的一般形式保持不变，平方和对于单元频率就保持不变。因此，我们常认为此类平方和对于不带缺失单元格的不平衡模型有用。在不带缺失单元的因子设计中，此方法等同于 Yates 加权均值平方方法。类型 III 平方和法常用于：

- 任何在类型 I 和类型 II 中列出的模型。
- 任何不带空白单元的平衡或非平衡模型。

类型 IV。此方法针对存在缺失单元的情况设计。对于设计中的任何效应 F，如果任何其它效应中不包含 F，则类型 IV = 类型 III = 类型 II。当 F 包含在其它效应中时，则类型 IV 将 F 中的参数中正在进行的对比相等地分配到所有较高水平的效应。类型 IV 平方和法常用于：

- 任何在类型 I 和类型 II 中列出的模型。
- 任何带有空白单元的平衡或非平衡模型。

GLM 多变量：对比

图片 2-3
“多变量：对比”对话框

“对比”用于检验效应水平之间是否存在显著性差异。您可以为模型中的每个因子指定一个对比。对比代表参数的线性组合。

假设检验基于原假设 LBM = 0，其中 L 是对比系数矩阵，M 是恒等矩阵，该矩阵的维数等于因变量的个数，B 是参数矢量。当指定对比之后，创建一个 L 矩阵，使得与因子对应的列与对比匹配。对剩余的列进行调整，使 L 矩阵可以估计。

除了使用 F 统计量的单变量检验和基于跨所有因变量的对比差分的 Student 的 t 分布的 Bonferroni 型同时置信区间以外，还提供使用 Pillai 的轨迹、Wilks 的 lambda、Hotelling 的轨迹以及 Roy 的最大根条件的多变量检验。

可用对比有偏移对比、简单对比、差分对比、Helmert 对比、重复对比和多项式对比。对于偏移对比和简单对比，您可以选择参考类别是最后一个类别还是第一个类别。

对比类型

偏差。将每个水平（参考类别除外）的均值与所有水平的均值（总均值）进行比较。因子的水平可以为任何顺序。

简单。将每个水平的均值与指定水平的均值进行比较。当存在控制组时，此类对比很有用。可以选择第一个或最后一个类别作为参考类别。

差分。将每个水平的均值（第一个水平除外）与前面水平的均值进行比较。（有时候称为逆 Helmert 对比。）

Helmert。将因子的每个水平的均值（最后一个水平除外）与后面水平的均值进行比较。

重复。将每个水平的均值（最后一个水平除外）与后一个水平的均值进行比较。

多项式。比较线性作用、二次作用、三次作用等等。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。

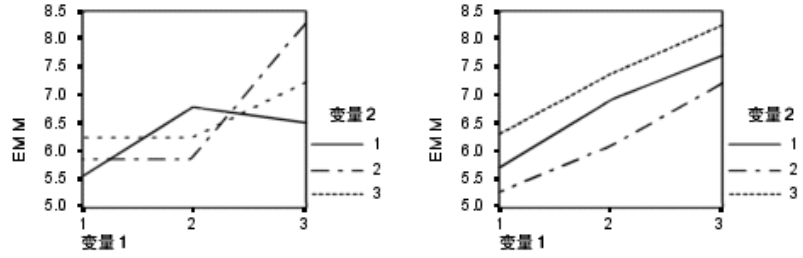
GLM 多变量：轮廓图

图片 2-4
“多变量：轮廓图”对话框

轮廓图（交互图）对于比较模型中的边际均值是有用的。轮廓图是一个线图，其中每个点表示因子的一个水平上的估计因变量边际均值（已针对任何协变量进行调整）。第二个因子的水平可用来绘制分离线。第三个因子中的每个水平可用来创建分离图。所有因子都可用于图。为每个因变量创建轮廓图。

单因子的轮廓图显示估计边际均值是沿水平增加还是减小。对于两个或更多因子，平行线表示因子之间没有交互，这意味着您只能调查一个因子的水平。不平行的线则表示交互。

图片 2-5
不平行图（左）和平行图（右）



在通过为水平轴选择因子，以及通过为分离线和分离图选择因子（后者可选）指定了图之后，该图必须添加到“图”列表中。

GLM 多变量：两两比较

图片 2-6
“多变量：观察到的均值的两两比较”对话框

两两比较检验。 一旦确定均值间存在差值，两两范围检验和成对多重比较就可以确定哪些均值存在差值了。对未调整的值进行比较。单独为每个因变量执行两两比较检验。

Bonferroni 和 Tukey’s 真实显著性差异检验是常用的多重比较检验。**Bonferroni 检验**基于 Student 的 t 统计量，它针对已进行多重比较这一事实调整观察的显著性水平。**Sidak 的 t 检验**也调整显著性水平，并提供比 Bonferroni 检验更严密的界限。**Tukey’s 真实显著性差异检验**使用 Student 化的范围统计量在组之间进行所有成对比较，并将试验误差率设置为所有成对比较的集合的误差率。当检验大量均值对时，Tukey’s 真实显著性差异检验比 Bonferroni 检验更有效。对于少量的对，Bonferroni 更有效。

Hochberg’s GT2 类似于 Tukey’s 真实显著性差异检验，但使用了 Student 化的最大值模数。通常 Tukey 的检验更有效。**Gabriel 的成对比较检验**也使用 Student 化的最大值模数，在单元格尺寸不等的情况下通常比 Hochberg’s GT2 更有效。当单元大小变化过大时，Gabriel 检验可能会变得随意。

Dunnett 的成对多重比较 t 检验将一组处理与单个控制均值进行比较。最后一个类别是缺省的控制类别。另外，您还可以选择第一个类别。您还可以选择双侧或单侧检验。要检验因子的任何水平（控制类别除外）的均值是否不等于控制类别的均值，请使用双侧检验。要检验因子的任何水平的均值是否小于控制类别的均值，请选择 < **控制**。类似地，要检验因子的任何水平的均值是否大于控制类别的均值，请选择 > **控制**。

Ryan、Einot、Gabriel 和 Welsch (R-E-G-W) 开发了两个多重逐步降低范围检验。多重逐步降低过程首先检验所有均值是否相等。如果不是所有的均值均相等，则检验一部分均值的相等性。**R-E-G-W F** 基于 F 检验，而 **R-E-G-W Q** 基于 Student 化的范围。这些检验要比 Duncan 的多范围检验和 Student-Newman-Keuls（也是多重逐步下降过程）有效，但对于不相等的单元大小则不推荐使用它们。

当方差不等时，使用 **Tamhane’s T2**（基于 t 检验的保守成对比较检验）、**Dunnett’s T3**（基于 Student 化的最大模数的成对比较检验）、**Games-Howell 成对比较检验**（有时是随意的）或者 **Dunnett’s C**（基于 Student 化的范围的成对比较检验）。

Duncan 的多范围检验、Student-Newman-Keuls（**S-N-K**）和 **Tukey 的 b** 是排列组均值等级的范围检验，并计算范围值。这些检验的使用频率不如先前讨论的检验。

Waller-Duncan t 检验使用 Bayesian 方法。当样本大小不相等时，此范围检验使用样本大小的调和均值。

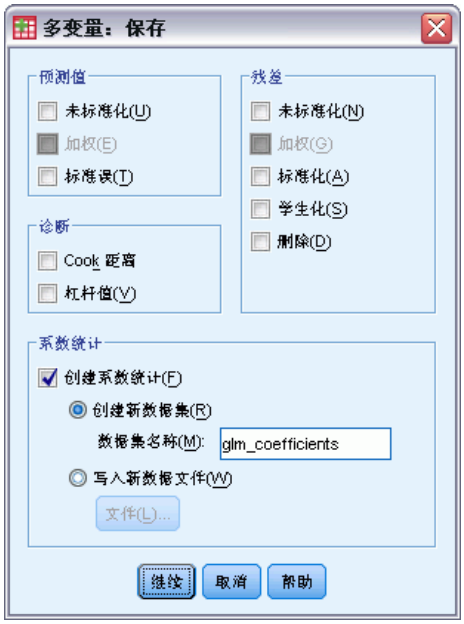
Scheffé 检验的显著性水平可允许要检验的组均值的所有可能的线性组合，而不仅仅是此功能中可用的成对比较。其结果是，Scheffé 检验常常比其他检验更保守，这意味着对于显著性，需要均值之间有更大的差别。

最小显著性差异（**LSD**）成对多重比较检验等同于所有组对之间的多重个别 t 检验。此检验的缺点是，不进行任何尝试来为多重比较调整观察到的显著性水平。

显示的检验。为 LSD、Sidak、Bonferroni、Games-Howell、Tamhane’s T2 和 T3、Dunnett’s C 以及 Dunnett’s T3 提供成对比较。为 S-N-K、Tukey 的 b、Duncan、R-E-G-W F、R-E-G-W Q 以及 Waller 提供范围检验的均一子集。Tukey’s 真实显著性差异检验、Hochberg’s GT2、Gabriel 的检验以及 Scheffé 的检验既是多重比较检验，同时也是多范围检验。

GLM: 保存

图片 2-7
保存对话框



您可以在数据编辑器中将模型预测的值、残差和相关测量另存为新变量。这些变量中有许多可用于检查关于数据的假设。要保存供另一 IBM® SPSS® Statistics 会话中使用的值，您必须保存当前数据文件。

预测值。模型为每个个案预测的值。

- **未标准化.** 模型为因变量预测的值。

- **加权**. 加权未标准化预测值。仅在之前已选择了 WLS 变量的情况下可用。
- **标准误**. 对于自变量具有相同值的个案所对应的因变量均值标准差的估计。

诊断。标识以下个案的测量：自变量的值具有不寻常组合的个案，以及可能对模型产生很大影响的个案。

- **Cook 距离**. 在特定个案从回归系数的计算中排除的情况下，所有个案的残差变化幅度的测量。较大的 Cook 距离表明从回归统计量的计算中排除个案之后，系数会发生根本变化。
- **杠杆值**. 未居中的杠杆值。每个观察值对模型拟合的相对影响。

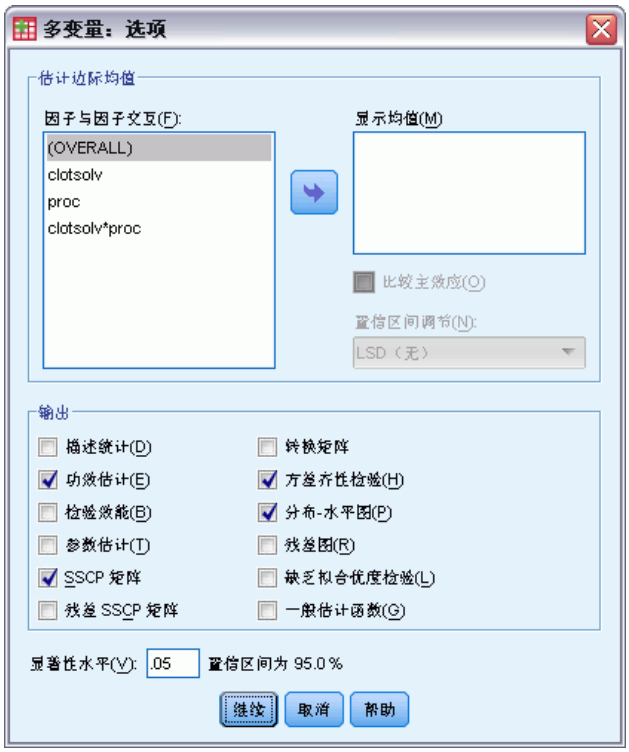
残差。未标准化残差是因变量的实际值减去由模型预测的值。还提供标准化残差、Student 化的残差以及剔除残差。如果选择了 WLS 变量，则提供加权的未标准化残差。

- **未标准化**. 观察值与模型预测值之间的差。
- **加权**. 加权未标准化残差。仅在之前已选择了 WLS 变量的情况下可用。
- **标准化**. 残差除以其标准差的估计。标准化残差也称为 Pearson 残差，它的均值为 0，标准差为 1。
- **学生化**. 残差除以其随个案变化的标准差的估计，这取决于每个个案的自变量值与自变量均值之间的距离。
- **删除**. 当某个案从回归系数的计算中排除时，该个案的残差。它是因变量的值和调整预测值之间的差。

系数统计。将模型中的参数估计值的协方差矩阵写入当前会话中的新数据集，或写入外部 SPSS Statistics 数据文件。而且，对于每个因变量，将存在一行参数估计值、一行与参数估计值对应的 t 统计量的显著性值以及一行残差自由度。对于多变量模型，每一个因变量都存在类似的行。您可以在读取矩阵文件的其他过程中使用此矩阵文件。

GLM 多变量：选项

图片 2-8
“多变量：选项”对话框



此对话框中有一些可选统计量。统计量是使用固定效应模型计算的。

估计边际均值。选择您需要的单元中的总体边际均值估计的因子和交互作用。为协变量（如果存在）调整这些均值。仅当指定了定制模型时交互才可用。

- **比较主效应。**对于主体间和主体内因子，为模型中的任何主效应提供估计边际均值未修正的成对比较。只有在“显示以下项的均值”列表中选择了主效应的情况下，此项才可用。
- **置信区间调节。**选择最小显著性差异 (LSD)、Bonferroni 或对置信区间和显著性的 Sidak 调整。此项只有在选择了**比较主作用**的情况下才可用。

输出。选择**描述统计**以生成所有单元中的所有因变量的观察到的均值、标准差和计数。**功效估计**给出了每个作用和每个参数估计值的偏 eta 方值。eta 方统计量描述总变异性中可归因于某个因子的部分。当基于观察到的值设置备用假设时，选择**检验效能**可获取检验的效能。选择**参数估计**可为每个检验生成参数估计值、标准误、t 检验、置信区间和检验效能。可以显示假设和误差 **SSCP 矩阵**以及**残差 SSCP 矩阵**加上残差协方差矩阵的 Bartlett 球形度检验。

齐性检验为跨主体间因子所有水平组合的每个因变量生成 Levene 的方差齐性检验（仅对于主体间因子）。另外，齐性检验包含对因变量协方差矩阵的齐性 Box M 检验，这些因变量跨主体间因子的所有水平组合。分布-水平图和残差图选项对于检查关于数据的假设很有用。如果不存在任何因子，则禁用此项。选择**残差图**为每个因变量生成观察-预测-标准化残差图。这些图对于调查方差相等的假设很有用。选择**缺乏拟合优度检验**以检查因变

量和自变量之间的关系是否能由模型充分地描述。**常规可估计函数**允许您基于常规可估计函数构造定制的假设检验。任何对比系数矩阵中的行均是常规可估计函数的线性组合。

显著性水平。您可能想要调整用在两两比较检验中的显著性水平，以及用于构造置信区间的置信度。指定的值还用于计算检验的检验效能。如果指定了显著性水平，则相关联的置信区间度会显示在对话框中。

GLM 命令的附加功能

这些功能可以适用于单变量、多变量或重复测量分析。使用命令语法语言还可以：

- 在设计中指定嵌套效应（使用 DESIGN 子命令）。
- 指定效应对比效应的线性组合或一个值的检验（使用 TEST 子命令）。
- 指定多个对比（使用 CONTRAST 子命令）。
- 包括用户缺失值（使用 MISSING 子命令）。
- 指定 EPS 标准（使用 CRITERIA 子命令）。
- 构造定制的 L 矩阵、M 矩阵或 K 矩阵（使用 LMATRIX、MMATRIX 或 KMATRIX 子命令）。
- 为偏移对比或简单对比指定中间参考类别（使用 CONTRAST 子命令）。
- 为多项式对比指定矩阵（使用 CONTRAST 子命令）。
- 为两两比较指定误差项（使用 POSTHOC 子命令）。
- 为因子列表中的任何因子或因子之间的因子交互计算估计边际均值（使用 EMMEANS 子命令）。
- 为临时变量指定名称（使用 SAVE 子命令）。
- 构造相关矩阵数据文件（使用 OUTFILE 子命令）。
- 构造包含主体间 ANOVA 表中的统计量的矩阵数据文件（使用 OUTFILE 子命令）。
- 将设计矩阵保存到新的数据文件（使用 OUTFILE 子命令）。

请参见命令语法参考以获取完整的语法信息。

章
3

GLM 重复测量

“GLM 重复测量”过程在对每个主体或个案多次执行相同的测量时提供方差分析。如果指定了主体间因子，这些因子会将总体划分成组。通过使用此一般线性模型过程，您可以检验关于主体间因子和主体内因子的效应的原假设。可以调查因子之间的交互以及单个因子的效应。另外，还可以包含常数协变量的效应以及协变量与主体间因子的交互。

在双重多变量重复测量设计中，因变量表示主体内因子不同水平的多个变量的测量。例如，您可能在三个不同的时间对每个主体同时测量了脉搏和呼吸。

“GLM 重复测量”过程提供了对重复测量数据的单变量和多变量分析。平衡与非平衡模型均可进行检验。如果模型中的每个单元包含相同的个案数，则设计是平衡的。在多变量模型中，模型中的效应引起的平方和以及误差平方和以矩阵形式表示，而不是以单变量分析中的标量形式表示。这些矩阵称为 SSCP（平方和与叉积）矩阵。除了检验假设，“GLM 重复测量”过程还生成参数估计。

常用的先验对比可用于对主体间因子执行假设检验。另外，在整体的 F 检验已显示显著性之后，可以使用两两比较检验评估指定均值之间的差值。估计边际均值为模型中的单元提供了预测均值估计值，且这些均值的轮廓图（交互图）允许您轻松对其中一些关系进行可视化。

残差、预测值、Cook 距离以及杠杆值可以另存为数据文件中检查假设的新变量。另外还提供残差 SSCP 矩阵（残差的平方和与叉积的方形矩阵）、残差协方差矩阵（残差 SSCP 矩阵除以残差的自由度）和残差相关矩阵（残差协方差矩阵的标准化形式）。

WLS 权重允许您指定一个变量，用来针对加权最小平方 (WLS) 分析为观察值赋予不同权重，这样也许可以补偿测量的不同精确度。

示例。根据学生的焦虑程度检验的得分将十二个学生分配到高或低焦虑程度组。焦虑等级被认为是主体间因子，因为它会将主体划分成组。让每个学生进行四个学习任务试验，并记录每次试验中所犯错误的个数。每次试验的错误都记录在单独的变量中，并使用四个试验的四个水平定义主体内因子（试验）。试验的效果很明显，而试验与焦虑的交互则不明显。

方法。类型 I、类型 II、类型 III 和类型 IV 的平方和可用来评估不同的假设。类型 III 是缺省值。

统计量。两两比较范围检验和多重比较（对于主体间因子）：最小显著性差异、Bonferroni、Sidak、Scheffé、Ryan-Einot-Gabriel-Welsch 多重 F、Ryan-Einot-Gabriel-Welsch 多范围、Student-Newman-Keuls、Tukey’s 真实显著性差异、Tukey 的 b、Duncan、Hochberg’s GT2、Gabriel、Waller-Duncan t 检验、Dunnett（单侧和双侧）、Tamhane’s T2、Dunnett’s T3、Games-Howell 和 Dunnett’s C。描述统计：所有单元中所有因变量的观察均值、标准差和计数；Levene 的方差齐性检验；对因变量协方差矩阵的齐性 Box 的 M 检验以及 Mauchly 球形度检验。

图。分布-水平图、残差图以及轮廓图（交互）。

数据。因变量应是定量的。主体间因子将样本划分为离散的子组，例如男性和女性。这些因子应是分类因子，可以具有数字值或字符串值。主体内因子是在“重复测量定义因子”对话框中定义的。协变量是与因变量相关的定量变量。对于重复测量分析，这些数据在每个主体内变量水平都应该保持不变。

数据文件中应该为主体的每组测量包含一组变量。该组变量为组中的每次重复测量包含一个变量。为水平数等于重复次数的组定义一个主体内因子。例如，进行权重测量可能需要不同的天数。如果在五天内测量相同的属性，则主体内因子可以指定为 day，并且该因子具有五个水平。

对于多个主体内因子，每个主体的测量次数均等于每个因子的水平数的乘积。例如，如果四天内在每天的三个不同时间进行测量，则每个主体的总测量次数为 12。主体内因子可指定为 day(4) 和 time(3)。

假设。重复测量分析可通过两种方式完成，即单变量和多变量。

单变量方法（也称为分割图或混合模型方法）将因变量视为对主体内因子的水平的响应。主体测量应为来自多变量正态分布的样本，方差-协方差矩阵在主体间效应形成的单元内应该都相同。有些假设是针对因变量的方差-协方差矩阵的。如果方差-协方差矩阵是圆形的，单变量方法中使用的 F 统计量的有效性就可以得到保证（Huynh and Mandeville，1979 年）。

要检验此假设，可以使用 Mauchly 球形度检验，该方法会对进行了正交标准化转换的因变量的方差-协方差矩阵执行球形度检验。对于重复测量分析，Mauchly 检验会自动显示。对于较小的样本，此检验表现的功能并不十分强大。对于较大的样本，此检验的效果可能显而易见，即使是在偏差对结果的影响很小的情况下也不例外。如果检验的显著性很大，则可采用球形度假设。不过，在显著性不大并且似乎违反了球形度假设的情况下，可以对分子和分母自由度进行一定的调整，以便验证单变量 F 统计量。“GLM 重复测量”过程中存在三个对此调整的估计值，称为 **epsilon**。分子和分母自由度都必须乘以 epsilon，并使用新的自由度估计 F 比的显著性。

多变量方法将主体测量视为来自多变量正态分布的样本，方差-协方差矩阵在主体间效应形成的单元内应该都相同。要检验方差-协方差矩阵是否在所有单元内都相同，可以使用 Box M 检验。

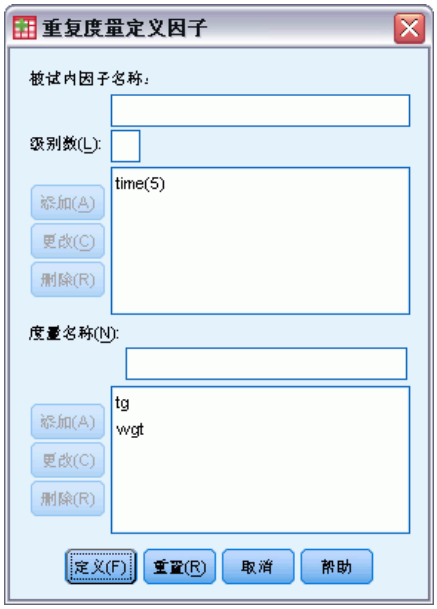
相关过程。在进行方差分析之前使用“探索”过程来检查数据。如果不 存在对每个主体的重复测量，则请使用“GLM 单变量”或“GLM 多变量”。如果每个主体仅存在两个测量（例如检验前和检验后测量），并且不存在主体间因子，则可以使用“配对样本 T 检验”过程。

获取 GLM 重复测量

▶ 从菜单中选择：

分析 > 一般线性模型 > 重复测量...

图片 3-1
“重复测量定义因子”对话框



- 群体内部因子名称及其级别数。
- 单击**添加**。
- 对每个主体内因子重复这些步骤。

为双重多变量重复测量设计定义测量因子：

- 键入测量名称。
- 单击**添加**。

在定义完所有因子和测量后：

- 单击**定义**。

图片 3-2
“重复测量”对话框

► 在列表上选择一个与群体内部因子（也可能是测量）的每个组合对应的因变量。

要更改变量位置，请使用向上和向下箭头。

要更改群体内部因子，可在不关闭主对话框的情况下重新打开“重复测量定义因子”对话框。或者，您也可以指定主体间因子和协变量。

GLM 重复测量定义因子

“GLM 重复测量”可分析表示相同属性的不同测量的相关因变量的分组。此对话框使您可以定义一个或多个主体内因子以便在“GLM 重复测量”中使用。请参阅第 15 页码中的图片 3-1. 注意，指定主体内因子的顺序非常重要。每个因子都构成前一个因子内的一个水平。

要使用“重复测量”，必须正确设置数据。您必须在此对话框中定义主体内因子。注意，这些因子不是数据内的现有变量，而是在这里定义的因子。

示例。在减肥研究中，假设要连续五周每周测量几个人的体重。在数据文件中，每个人就是一个主体或者个案。每周测量到的体重都记录在 weight1、weight2 等变量中。每个人的性别则记录在其他变量中。通过定义主体内因子可以将为每个主体重复测量的体重分组。该因子可命名为 week，并将其定义为具有五个水平。在主对话框中，变量 weight1 到 weight5 用于指定 week 的五个水平。数据文件中将女性和男性分组的变量 (gender) 可以指定为主体间因子以便研究男性和女性的差别。

测量。如果每次就多个测量对主体进行检验，请定义测量。例如，可以在一周的每一天对每个主体的脉搏和呼吸比率进行测量。这些测量不作为变量保存在数据文件中，但是在此处定义。具有多个测量的模型有时称为双重多变量重复测量模型。

GLM 重复测量：模型

图片 3-3
“重复测量：模型”对话框

指定模型。全因子模型包含所有因子主效应、所有协变量主效应以及所有因子间交互。它不包含协变量交互。选择**定制**可以仅指定其中一部分的交互或指定因子协变量交互。必须指定要包含在模型中的所有项。

主体间。列出主体间因子与协变量。

模型。模型取决于数据的性质。选择**定制**后，您可以选择分析中感兴趣的主体内效应和交互以及主体间效应和交互。

平方和。计算主体间模型的平方和的方法。对于没有缺失单元的平衡或非平衡主体间模型，类型 III 平方和法最常用。

构建项

对于选定因子和协变量：

交互。创建所有选定变量的最高级交互项。这是缺省值。

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二阶交互。

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。

平方和

对于该模型，您可以选择平方和类型。类型 III 最常用，并且是缺省类型。

类型 I。此方法也称为平方和分级解构法。在模型中，每一项只针对它前面的那项进行调整。类型 I 平方和常用于：

- 平衡 ANOVA 模型，其中任何主效应在任何一阶交互效应之前指定，任何一阶交互效应在任何二阶交互效应之前指定，依此类推。
- 多项式回归模型，其中任何低阶项在任何高阶项之前指定。
- 纯嵌套模型，其中第一个指定的效应嵌套在第二个指定的效应中，第二个指定的效应嵌套在第三个指定的效应中，依此类推。（此嵌套形式只能通过使用语法来指定。）

类型 II。此方法在为所有其它“相应的”效应进行调节的模型中计算某个效应的平方和。相应的效应是指，与所有效应（不包含正被检查的效应）相对应的效应。类型 II 平方和方法常用于：

- 平衡 ANOVA 模型。
- 任何只有主要因子效应的模型。
- 任何回归模型。
- 纯嵌套设计。（此嵌套形式能通过使用语法来指定。）

类型 III。缺省类型。此方法在设计中通过以下形式计算某个效应的平方和：为任何不包含该效应的其他效应，以及任何与包含该效应正交的效应（如果存在）调整的平方和。类型 III 平方和具有一个主要优点，那就是只要可估计性的一般形式保持不变，平方和对于单元频率就保持不变。因此，我们常认为此类平方和对于不带缺失单元格的不平衡模型有用。在不带缺失单元的因子设计中，此方法等同于 Yates 加权均值平方方法。类型 III 平方和法常用于：

- 任何在类型 I 和类型 II 中列出的模型。
- 任何不带空白单元的平衡或非平衡模型。

类型 IV。此方法针对存在缺失单元的情况设计。对于设计中的任何效应 F，如果任何其它效应中不包含 F，则类型 IV = 类型 III = 类型 II。当 F 包含在其它效应中时，则类型 IV 将 F 中的参数中正在进行的对比相等地分配到所有较高水平的效应。类型 IV 平方和法常用于：

- 任何在类型 I 和类型 II 中列出的模型。
- 任何带有空白单元的平衡或非平衡模型。

GLM 重复测量：对比

图片 3-4
“重复测量：对比”对话框

使用对比可以检验主体间因子的水平之间的差别。您可以为模型中的每个主体间因子指定一个对比。对比代表参数的线性组合。

假设检验基于原假设 LBM = 0，其中 L 是对比系数矩阵，B 是参数矢量，M 是对应于因变量的平均转换的平均矩阵。可以通过选择“重复测量：选项”对话框中的**转换矩阵**显示此转换矩阵。例如，如果有四个因变量和四个水平的主体内因子，且多项式对比（缺省值）用于主体内因子，则 M 矩阵将为 (0.5 0.5 0.5 0.5)’。当指定对比之后，创建一个 L 矩阵，使得与主体间因子对应的列与对比匹配。对剩余的列进行调整，使 L 矩阵可以估计。

可用对比有偏移对比、简单对比、差分对比、Helmert 对比、重复对比和多项式对比。对于偏移对比和简单对比，您可以选择参考类别是最后一个类别还是第一个类别。

必须为主体内因子选择一个除**无**之外的对比系数。

对比类型

偏差。将每个水平（参考类别除外）的均值与所有水平的均值（总均值）进行比较。因子的水平可以为任何顺序。

简单。将每个水平的均值与指定水平的均值进行比较。当存在控制组时，此类对比很有用。可以选择第一个或最后一个类别作为参考类别。

差分。将每个水平的均值（第一个水平除外）与前面水平的均值进行比较。（有时候称为逆 Helmert 对比。）

Helmert。将因子的每个水平的均值（最后一个水平除外）与后面水平的均值进行比较。

重复。将每个水平的均值（最后一个水平除外）与后一个水平的均值进行比较。

多项式。比较线性作用、二次作用、三次作用等等。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。

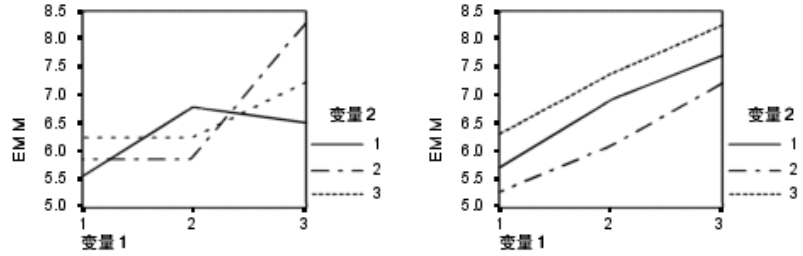
GLM 重复测量：轮廓图

图片 3-5
“重复测量：轮廓图”对话框

轮廓图（交互图）对于比较模型中的边际均值是有用的。轮廓图是一个线图，其中每个点表示因子的一个水平上的估计因变量边际均值（已针对任何协变量进行调整）。第二个因子的水平可用来绘制分离线。第三个因子中的每个水平可用来创建分离图。所有因子都可用于图。为每个因变量创建轮廓图。主体间因子和主体内因子都可用在轮廓图中。

单因子的轮廓图显示估计边际均值是沿水平增加还是减小。对于两个或更多因子，平行线表示因子之间没有交互，这意味着您只能调查一个因子的水平。不平行的线则表示交互。

图片 3-6
不平行图（左）和平行图（右）



在通过为水平轴选择因子，以及通过为分离线和分离图选择因子（后者可选）指定了图之后，该图必须添加到“图”列表中。

GLM 重复测量：两两比较

图片 3-7
“重复测量：观察到的均值的两两比较”对话框

两两比较检验。 一旦确定均值间存在差值，两两范围检验和成对多重比较就可以确定哪些均值存在差值了。对未调整的值进行比较。如果没有主体间因子，则这些检验不可用，并为跨主体内因子的水平的均值执行两两比较检验。

Bonferroni 和 Tukey’s 真实显著性差异检验是常用的多重比较检验。**Bonferroni 检验**基于 Student 的 t 统计量，它针对已进行多重比较这一事实调整观察的显著性水平。**Sidak 的 t 检验**也调整显著性水平，并提供比 Bonferroni 检验更严密的界限。**Tukey’s 真实显著性差异检验**使用 Student 化的范围统计量在组之间进行所有成对比较，并将试验误差率设置为所有成对比较的集合的误差率。当检验大量均值对时，Tukey’s 真实显著性差异检验比 Bonferroni 检验更有效。对于少量的对，Bonferroni 更有效。

Hochberg’s GT2 类似于 Tukey扭 真实显著性差异检验，但使用了 Student 化的最大值模数。通常 Tukey 的检验更有效。**Gabriel 的成对比较检验**也使用 Student 化的最大值模数，在单元格尺寸不等的情况下通常比 Hochberg’s GT2 更有效。当单元大小变化过大时，Gabriel 检验可能会变得随意。

Dunnett 的成对多重比较 t 检验将一组处理与单个控制均值进行比较。最后一个类别是缺省的控制类别。另外，您还可以选择第一个类别。您还可以选择双侧或单侧检验。要检验因子的任何水平（控制类别除外）的均值是否不等于控制类别的均值，请使用双侧检验。要检验因子的任何水平的均值是否小于控制类别的均值，请选择 < **控制**。类似地，要检验因子的任何水平的均值是否大于控制类别的均值，请选择 > **控制**。

Ryan、Einot、Gabriel 和 Welsch (R-E-G-W) 开发了两个多重逐步降低范围检验。多重逐步降低过程首先检验所有均值是否相等。如果不是所有的均值均相等，则检验一部分均值的相等性。**R-E-G-W F** 基于 F 检验，而 **R-E-G-W Q** 基于 Student 化的范围。这些检验要比 Duncan 的多范围检验和 Student-Newman-Keuls（也是多重逐步下降过程）有效，但对于不相等的单元大小则不推荐使用它们。

当方差不等时，使用 **Tamhane’s T2**（基于 t 检验的保守成对比较检验）、**Dunnett’s T3**（基于 Student 化的最大模数的成对比较检验）、**Games-Howell 成对比较检验**（有时是随意的）或者 **Dunnett’s C**（基于 Student 化的范围的成对比较检验）。

Duncan 的多范围检验、Student-Newman-Keuls（**S-N-K**）和 **Tukey 的 b** 是排列组均值等级的范围检验，并计算范围值。这些检验的使用频率不如先前讨论的检验。

Waller-Duncan t 检验使用 Bayesian 方法。当样本大小不相等时，此范围检验使用样本大小的调和均值。

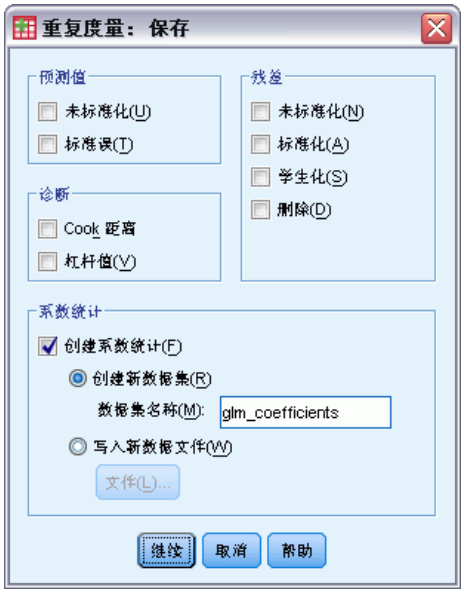
Scheffé 检验的显著性水平可允许要检验的组均值的所有可能的线性组合，而不仅仅是此功能中可用的成对比较。其结果是，Scheffé 检验常常比其他检验更保守，这意味着对于显著性，需要均值之间有更大的差别。

最小显著性差异（**LSD**）成对多重比较检验等同于所有组对之间的多重个别 t 检验。此检验的缺点是，不进行任何尝试来为多重比较调整观察到的显著性水平。

显示的检验。为 LSD、Sidak、Bonferroni、Games-Howell、Tamhane’s T2 和 T3、Dunnett’s C 以及 Dunnett’s T3 提供成对比较。为 S-N-K、Tukey 的 b、Duncan、R-E-G-W F、R-E-G-W Q 以及 Waller 提供范围检验的均一子集。Tukey扭 真实显著性差异检验、Hochberg’s GT2、Gabriel 的检验以及 Scheffé 的检验既是多重比较检验，同时也是多范围检验。

GLM 重复测量：保存

图片 3-8
“重复测量：保存”对话框



您可以在数据编辑器中将模型预测的值、残差和相关测量另存为新变量。这些变量中有许多可用于检查关于数据的假设。要保存供另一 IBM® SPSS® Statistics 会话中使用的值，您必须保存当前数据文件。

预测值。模型为每个个案预测的值。

- **未标准化**. 模型为因变量预测的值。
- **标准误**. 对于自变量具有相同值的个案所对应的因变量均值标准差的估计。

诊断。标识以下个案的测量：自变量的值具有不寻常组合的个案，以及可能对模型产生很大影响的个案。可用的有 Cook 距离以及不居中的杠杆值。

- **Cook 距离**. 在特定个案从回归系数的计算中排除的情况下，所有个案的残差变化幅度的测量。较大的 Cook 距离表明从回归统计量的计算中排除个案之后，系数会发生根本变化。
- **杠杆值**. 未居中的杠杆值。每个观察值对模型拟合的相对影响。

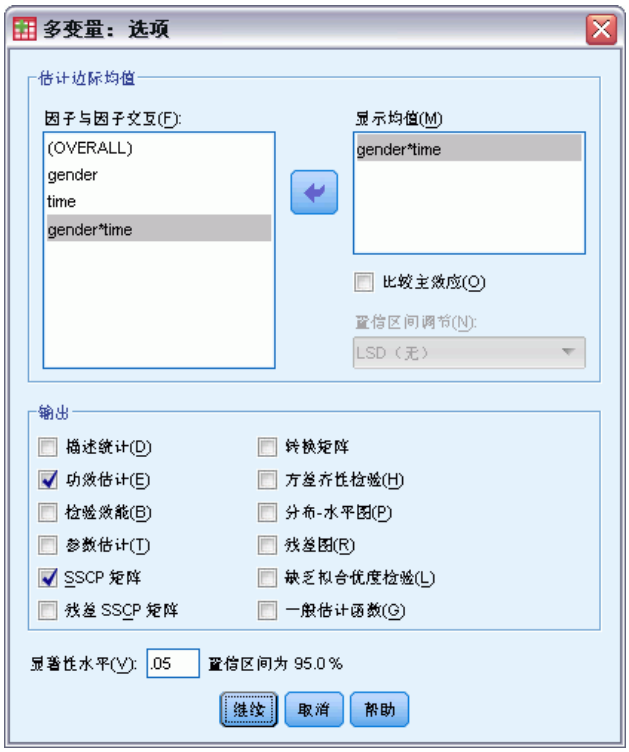
残差。未标准化残差是因变量的实际值减去由模型预测的值。还提供标准化残差、Student 化的残差以及剔除残差。

- **未标准化**. 观察值与模型预测值之间的差。
- **标准化**. 残差除以其标准差的估计。标准化残差也称为 Pearson 残差，它的均值为 0，标准差为 1。
- **学生化**. 残差除以其随个案变化的标准差的估计，这取决于每个个案的自变量值与自变量均值之间的距离。
- **删除**. 当某个案从回归系数的计算中排除时，该个案的残差。它是因变量的值和调整预测值之间的差。

系数统计。将参数估计值的方差-协方差矩阵保存到数据集或数据文件。而且，对于每个因变量，将存在一行参数估计值、一行与参数估计值对应的 t 统计量的显著性值以及一行残差自由度。对于多变量模型，每一个因变量都存在类似的行。您可以在读取矩阵文件的其他过程中使用此矩阵数据。可以在同一会话中继续使用数据集，但不会将其另存为文件，除非在会话结束之前明确将其保存为文件。数据集名称必须符合变量命名规则。

GLM 重复测量：选项

图片 3-9
“重复测量：选项”对话框



此对话框中有一些可选统计量。统计量是使用固定效应模型计算的。

估计边际均值。选择您需要的单元中的总体边际均值估计的因子和交互作用。为协变量（如果存在）调整这些均值。可以选择主体内因子和主体间因子。

- **比较主效应。**对于主体间和主体内因子，为模型中的任何主效应提供估计边际均值未修正的成对比较。只有在“显示以下项的均值”列表中选择了主效应的情况下，此项才可用。
- **置信区间调节。**选择最小显著性差异 (LSD)、Bonferroni 或对置信区间和显著性的 Sidak 调整。此项只有在选择了**比较主作用**的情况下才可用。

输出。选择**描述统计**以生成所有单元中的所有因变量的观察到的均值、标准差和计数。**功效估计**给出了每个作用和每个参数估计值的偏 eta 方值。eta 方统计量描述总变异性中可归因于某个因子的部分。当基于观察到的值设置备用假设时，选择**检验效能**可获取检验的效能。选择**参数估计**可为每个检验生成参数估计值、标准误、t 检验、置信区间和检验效能。可以显示假设和误差 **SSCP 矩阵**以及**残差 SSCP 矩阵**加上残差协方差矩阵的 Bartlett 球形度检验。

齐性检验为跨主体间因子所有水平组合的每个因变量生成 Levene 的方差齐性检验（仅对于主体间因子）。另外，齐性检验包含对因变量协方差矩阵的齐性 Box M 检验，这些因变量跨主体间因子的所有水平组合。分布-水平图和残差图选项对于检查关于数据的假设很有用。如果不存在任何因子，则禁用此项。选择**残差图**为每个因变量生成观察-预测-标准化残差图。这些图对于调查方差相等的假设很有用。选择**缺乏拟合优度检验**以检查因变

量和自变量之间的关系是否能由模型充分地描述。**常规可估计函数**允许您基于常规可估计函数构造定制的假设检验。任何对比系数矩阵中的行均是常规可估计函数的线性组合。

显著性水平。您可能想要调整用在两两比较检验中的显著性水平，以及用于构造置信区间的置信度。指定的值还用于计算检验的检验效能。如果指定了显著性水平，则相关联的置信区间度会显示在对话框中。

GLM 命令的附加功能

这些功能可以适用于单变量、多变量或重复测量分析。使用命令语法语言还可以：

- 在设计中指定嵌套效应（使用 DESIGN 子命令）。
- 指定效应对比效应的线性组合或一个值的检验（使用 TEST 子命令）。
- 指定多个对比（使用 CONTRAST 子命令）。
- 包括用户缺失值（使用 MISSING 子命令）。
- 指定 EPS 标准（使用 CRITERIA 子命令）。
- 构造定制的 L 矩阵、M 矩阵或 K 矩阵（使用 LMATRIX、MMATRIX 和 KMATRIX 子命令）。
- 为偏移对比或简单对比指定中间参考类别（使用 CONTRAST 子命令）。
- 为多项式对比指定矩阵（使用 CONTRAST 子命令）。
- 为两两比较指定误差项（使用 POSTHOC 子命令）。
- 为因子列表中的任何因子或因子之间的因子交互计算估计边际均值（使用 EMMEANS 子命令）。
- 为临时变量指定名称（使用 SAVE 子命令）。
- 构造相关矩阵数据文件（使用 OUTFILE 子命令）。
- 构造包含主体间 ANOVA 表中的统计量的矩阵数据文件（使用 OUTFILE 子命令）。
- 将设计矩阵保存到新的数据文件（使用 OUTFILE 子命令）。

请参见命令语法参考以获取完整的语法信息。

章
4

方差成分分析

对于混合效应模型，“方差成分”过程估计每种随机效应对因变量方差的贡献。此过程对于混合模型的分析尤其有趣，例如分割图、单变量重复度量以及随机区组设计。通过计算方差成分，可以确定减小方差时的重点关注对象。

有四种不同的方法可用来估计方差成分：最小范数二次无偏估计 (MINQUE)、方差分析 (ANOVA)、最大似然 (ML) 和受约束的最大似然 (REML)。不同的方法具有各种不同的指定可供使用。

所有方法的缺省输出都包含方差成分估计。如果使用 ML 方法或 REML 方法，则还会显示一个渐近协方差矩阵表。对于 ANOVA 方法，其他可用的输出包括 ANOVA 表和期望均方，对于 ML 和 REML 方法，其他可用的输出包括迭代历史记录。“方差成分”过程与“GLM 单变量”过程完全兼容。

WLS 权重允许您指定一个变量，用来针对加权分析为观察值赋予不同权重，这样也许可以补偿不同的测量精确度偏差。

示例。某一农业学校测量六个不同猪栏中的猪一个月的重量增加量。猪栏这个变量是具有六个水平的随机因子。（进行研究的六个猪栏是来自大的猪栏总体的随机样本。）调查发现重量增长的方差更大程度上归因于猪栏的不同而不是猪栏中的猪的不同。

数据。因变量是定量变量。因子是分类变量。它们可以具有数字值或最多 8 个字节的字符串值。至少必须有一个因子是随机的。也就是说，因子的水平必须是可能的水平的随机样本。协变量是与因变量相关的定量变量。

假设。所有方法均假设随机效应的模型参数均值为零，方差为有限常数，并且模型参数互不相关。来自不同随机效应的模型参数也不相关。

残差项的均值也为零，方差也为有限常数。它与任何随机效应的模型参数都不相关。来自不同观察值的残差项被认为是不相关的。

基于这些假设，来自某一随机因子的相同水平的观察值是相关的。这就使得方差成分模型与一般线性模型区分开来。

ANOVA 和 MINQUE 不需要正态假设。它们对于对正态假设的适度偏差来说是稳健的。

ML 和 REML 要求模型参数和残差项服从正态分布。

相关过程。在进行方差成分分析之前使用“探索”过程来检查数据。对于假设检验，使用“GLM 单变量”、“GLM 多变量”和“GLM 重复测量”。

获取方差成分表

▶ 从菜单中选择：

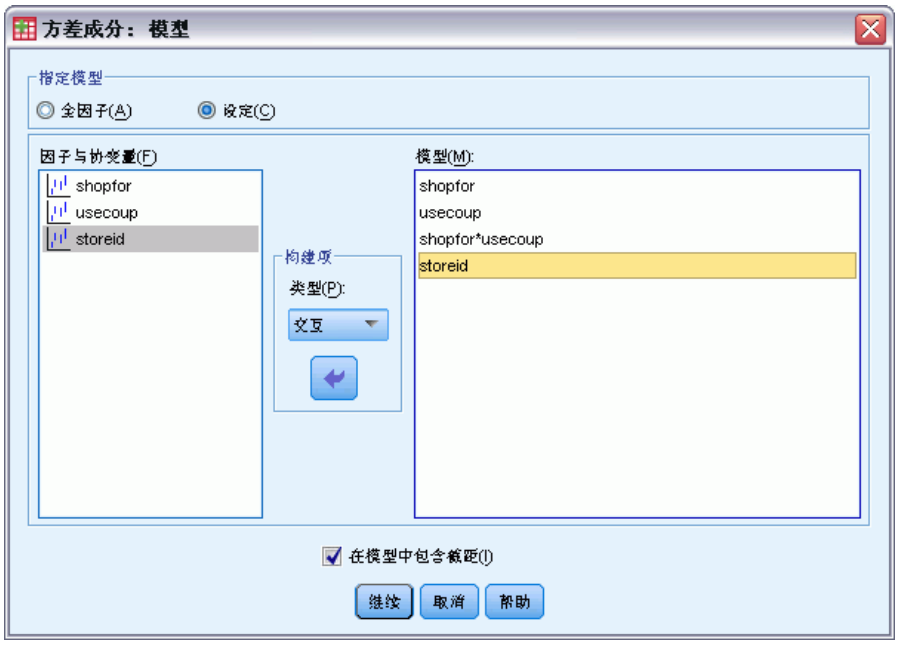
分析 > 一般线性模型 > 方差成分...

图片 4-1
“方差成分”对话框

- 选择一个因变量。
- 为“固定因子”、“随机因子”和“协变量”选择变量（如果适用于您的数据）。要指定权重变量，请使用“WLS 权重”。

方差成分：模型

图片 4-2
“方差成分：模型”对话框



指定模型。全因子模型包含所有因子主效应、所有协变量主效应以及所有因子间交互。它不包含协变量交互。选择**定制**可以仅指定其中一部分的交互或指定因子协变量交互。必须指定要包含在模型中的所有项。

因子与协变量。列出因子与协变量。

模型。模型取决于数据的性质。选择**定制**之后，您可以选择分析中感兴趣的主效应和交互效应。模型中必须包含随机因子。

在模型中包含截距。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。

构建项

对于选定因子和协变量：

交互。创建所有选定变量的最高级交互项。这是缺省值。

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二阶交互。

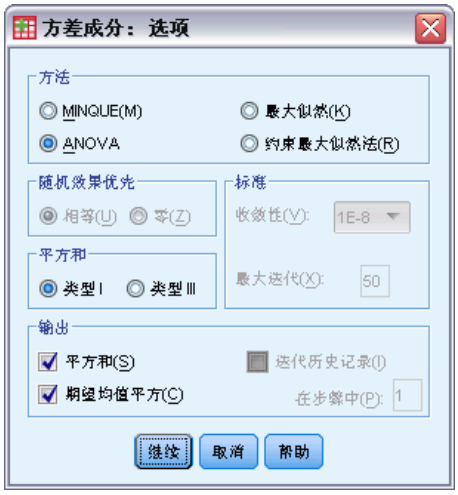
所有三阶。创建选定变量的所有可能的三阶交互。

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。

方差成分：选项

图片 4-3
“方差成分：选项”对话框



方法。您可以选择四种方法中的一种估计方差成分。

- MINQUE（最小范数二次无偏估计）可生成相对于固定效应不变的估计值。如果数据服从正态分布并且估计值是正确的，则此方法可生成所有无偏估计的最小方差。您可以为随机效应优先选择一种法。

- **ANOVA**（方差分析）使用每种效应的类型 I 或类型 III 平方和计算无偏估计。ANOVA 方法有时会生成负数方差估计，这可指示模型不正确、估计方法不合适或需要更多数据。
- **最大似然性** (ML) 使用迭代生成与实际观察到的数据最一致的估计值。这些估计值可能存在偏差。此方法是渐近正态分布。ML 和 REML 估计值在转换时保持不变。此方法不考虑估计固定效应时使用的自由度。
- **约束最大似然法** (REML) 估计在大多数（如果不是全部）平衡数据的情况下均可减少 ANOVA 估计值。由于此方法要针对固定效应进行调整，因此其标准误应比 ML 方法的标准误要小。此方法考虑估计固定效应时使用的自由度。

随机效果优先。统一意味着所有随机效应以及残差项对观察值具有相同的影响。**零**方案等同于假设随机效应方差为零。仅对 MINQUE 方法可用。

平方和。类型 I 平方和用于分层模型，分层模型常用于与方差成分有关的情况。如果选择 GLM 中的缺省选项**类型 III**，则方差估计值可用在“GLM 单变量”中，进行具有类型 III 平方和的假设检验。仅对 ANOVA 方法可用。

标准。您可以指定收敛标准和最大迭代次数。仅对 ML 或 REML 方法可用。

显示。对于 ANOVA 方法，您可以选择显示平方和与期望均值平方。如果选择了**最大似然性**或**约束最大似然法**，则可以显示迭代历史记录。

平方和（方差成分）

对于该模型，您可以选择一种平方和类型。类型 III 最常用，并且是缺省类型。

类型 I。此方法也称为平方和分级解构法。在模型中，每一项只针对它前面的那项进行调整。类型 I 平方和法常用于：

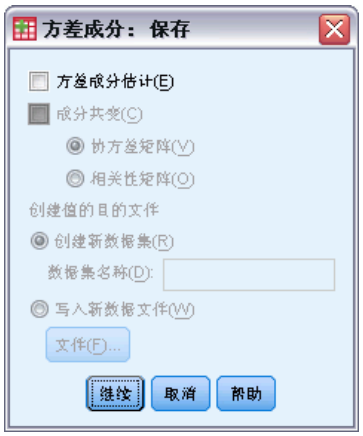
- 平衡 ANOVA 模型，其中任何主效应在任何一阶交互效应之前指定，任何一阶交互效应在任何二阶交互效应之前指定，依此类推。
- 多项式回归模型，其中任何低阶项在任何高阶项之前指定。
- 纯嵌套模型，其中第一个指定的效应嵌套在第二个指定的效应中，第二个指定的效应嵌套在第三个指定的效应中，依此类推。（此嵌套形式只能通过使用语法来指定。）

类型 III。缺省类型。此方法在设计中通过以下形式计算某个效应的平方和：为任何不包含该效应的其他效应，以及任何与包含该效应正交的效应（如果存在）调整的平方和。类型 III 平方和具有一个主要优点，那就是只要可估计性的一般形式保持不变，平方和对于单元频率就保持不变。因此，此类平方和常被认为对不带缺失单元的不平衡模型有用。在不带缺失单元的因子设计中，此方法等同于 Yates 加权均值平方方法。类型 III 平方和法常用于：

- “类型 I”中列出的所有模型。
- 任何不带空白单元的平衡或非平衡模型。

方差成分：保存到新文件

图片 4-4
“方差成分：保存到新文件”对话框



您可以将此过程的一些结果保存到新的 IBM® SPSS® Statistics 数据文件。

方差成分估计。将方差成分估计值和估计标签保存到数据文件或数据集。这些数据可用于计算更多统计量或 GLM 过程的进一步分析。例如，您可以使用这些数据计算置信区间或检验假设。

成分共变。将方差-协方差矩阵或相关矩阵保存到数据文件或数据集。仅当指定了**最大似然**或**受约束的最大似然**时才可用。

创建值的目的文件。允许您为包含方差成分估计值和/或矩阵的文件指定数据文件名称或外部文件名。可以在同一会话中继续使用数据集，但不会将其另存为文件，除非在会话结束之前明确将其保存为文件。数据集名称必须符合变量命名规则。

可以使用 MATRIX 命令从数据文件抽取需要的数据，然后计算置信区间或执行检验。

VARCOMP 命令的附加功能

使用命令语法语言还可以：

- 在设计中指定嵌套效应（使用 DESIGN 子命令）。
- 包括用户缺失值（使用 MISSING 子命令）。
- 指定 EPS 标准（使用 CRITERIA 子命令）。

请参见命令语法参考以获取完整的语法信息。

章
5

线性混合模型

“线性混合模型”过程扩展了一般线性模型，因此允许数据表现出相关的和不恒定的变异性。因此，线性混合模型提供了不仅能够就数据的均值还能够就其方差和协方差建模的灵活性。

此外，“线性混合模型”过程也是用于拟合可作为混合线性模型构建的其他模型的灵活工具。这些模型包括多变量模型、分层线性模型以及随机系数模型。

示例。有一家杂货连锁店想知道各种优惠券对客户消费的影响。通过抽取老客户的随机样本，他们记录了每个客户在过去 10 周内的消费情况。该公司每周向这些客户邮寄一种不同的优惠券。“线性混合模型”用于估计不同的优惠券对消费的影响，同时调整在 10 周内重复观察每个主体导致的相关性。

方法。最大似然 (ML) 和受约束的最大似然 (REML) 估计。

统计量。描述统计：各个不同的因子水平组合的因变量和协变量的样本大小、均值和标准差。因子水平信息：每个因子水平及其频率的排序值。此外，还有固定效应的参数估计值和置信区间，协方差矩阵的参数的 Wald 检验和置信区间。类型 I 和类型 III 的平方和可用于评估不同的假设。类型 III 是缺省值。

数据。因变量应是定量的。因子应是分类因子，可以具有数字值或字符串值。协变量和权重变量应是定量的。主体和重复变量可为任意类型。

假设。假设因变量与固定因子、随机因子和协变量线性相关。固定效应就因变量的均值建模。随机效应则就因变量的协方差结构建模。多个随机效应之间被认为是彼此独立的，并且会为每个效应计算一个单独的协方差矩阵；不过，针对同一随机效应指定的模型项可能是相关的。重复度量就残差的协方差结构建模。假定因变量也来自正态分布。

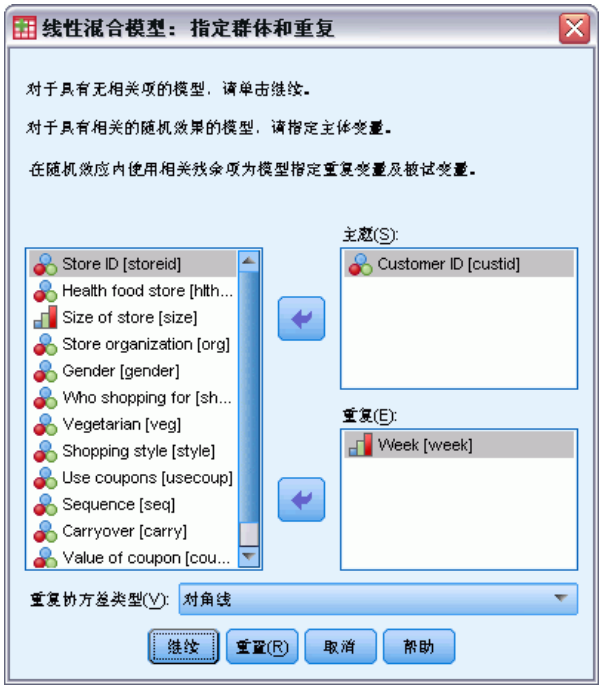
相关过程。在运行分析之前使用“探索”过程来检查数据。如果不怀疑相关的和不恒定的变异性的存在，则可改为使用“GLM 单变量”或“GLM 重复测量”过程。如果随机效应具有方差成分协方差结构，并且不存在重复度量，则可改用“方差成分分析”过程。

获取线性混合模型分析

▶ 从菜单中选择：

分析 > 混合模型 > 线性...

图片 5-1
“线性混合模型：指定主体和重复变量”对话框



- （可选）选择一个或多个群体变量。
- （可选）选择一个或多个重复变量。
- （可选）选择重复协方差类型。
- 单击继续。

图片 5-2
“线性混合模型”对话框

- 选择一个因变量。

- 至少选择一个因子或协变量。
- 单击**固定**或**随机**并至少指定一个固定效应或随机效应模型。

  根据需要，选择一个加权变量。

线性混合模型：选择主体/重复变量

此对话框允许选择定义主体和重复观察值的变量，也允许选择残差的协方差结构。 请参阅第 32 页码中的图片 5-1.

主体。主体是可视为独立于其他主体的观察单元。例如，在医学研究中可以认为某患者的血压读数独立于其他患者的读数。如果存在对每个主体的重复度量，而且您想要对这些观察值之间的相关性建模，定义主体就非常重要。例如，您可能期望同一个患者在连续多次就医时得到的血压读数是相关的。

主体也可由多个变量的因子水平组合进行定义；例如，您可以指定性别和年龄类别作为主体变量，就 males over the age of 65 相互类似但独立于 males under 65 和 females 的可信度进行建模。

主体列表中指定的所有变量都可用于定义残差协方差结构的主体。可以使用部分或者全部变量定义随机效应协方差结构的主体。

重复。在此列表中指定的变量用于标识重复观察值。例如，单个变量周可以标识医学研究中 10 周内的观察值，而月和天可共同用于标识一年内的每一天的观察值。

重复协方差类型。这指定残差的协方差结构。可用的结构如下：

- 前因:一阶。
- AR(1)
- AR(1):异质
- ARMA(1,1)
- 复合对称
- 复合对称：相关性度规。
- 复合对称：异质
- 对角线
- 因子分析：一阶。
- 因子分析：一阶、异质
- Huynh-Feldt
- 已标度的恒等
- Toeplitz
- Toeplitz：异质
- 未结构化
- 未结构化：相关

线性混合模型：固定效应

图片 5-3
“线性混合模型：固定效应”对话框

固定效应。没有缺省模型，因此必须显式指定固定效应。还可以构建嵌套或非嵌套项。

包括截距。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。

平方和。计算平方和的方法。对于不带缺失单元的模型，类型 III 方法最常用。

构建非嵌套项

对于选定因子和协变量：

因子。创建选定变量的所有可能的交互和主效应。这是缺省值。

交互。创建所有选定变量的最高级交互项。

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二阶交互。

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。

建立嵌套项

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客户的消费习惯。由于每位客户只频繁光顾某一位置的商店，因此客户效应**可以说是嵌套在**商店位置效应中。

此外，还可以包含交互效应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。

平方和

对于该模型，您可以选择平方和类型。类型 III 最常用，并且是缺省类型。

类型 I。此方法也称为平方和分级解构法。在模型中，每一项只针对它前面的那项进行调整。类型 I 平方和常用于：

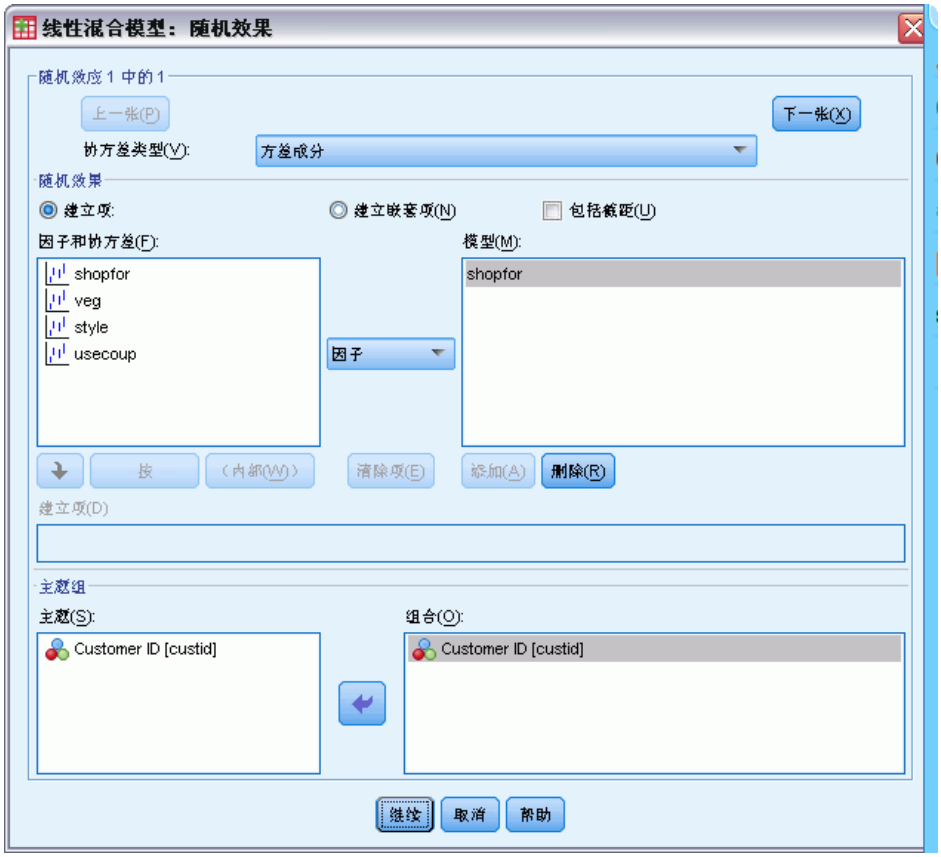
- 平衡 ANOVA 模型，其中任何主效应在任何一阶交互效应之前指定，任何一阶交互效应在任何二阶交互效应之前指定，依此类推。
- 多项式回归模型，其中任何低阶项在任何高阶项之前指定。
- 纯嵌套模型，其中第一个指定的效应嵌套在第二个指定的效应中，第二个指定的效应嵌套在第三个指定的效应中，依此类推。（此嵌套形式只能通过使用语法来指定。）

类型 III。缺省类型。此方法在设计中通过以下形式计算某个效应的平方和：为任何不包含该效应的其他效应，以及任何与包含该效应正交的效应（如果存在）调整的平方和。类型 III 平方和具有一个主要优点，那就是只要可估计性的一般形式保持不变，平方和对于单元频率就保持不变。因此，我们常认为此类平方和对于不带缺失单元格的不平衡模型有用。在不带缺失单元的因子设计中，此方法等同于 Yates 加权均值平方方法。类型 III 平方和法常用于：

- “类型 I”中列出的所有模型。
- 任何不带空白单元的平衡或非平衡模型。

线性混合模型：随机效果

图片 5-4
“线性混合模型：随机效果”对话框



协方差类型。允许您为随机效应模型指定协方差结构。会为每个随机效应估计一个单独的协方差矩阵。可用的结构如下：

- 前因:一阶。
- AR(1)
- AR(1):异质
- ARMA(1,1)
- 复合对称
- 复合对称：相关性度规。
- 复合对称：异质
- 对角线
- 因子分析：一阶。
- 因子分析：一阶、异质
- Huynh-Feldt
- 已标度的恒等

- Toeplitz
- Toeplitz：异质
- 未结构化
- 未结构化：相关性度规。
- Variance Components

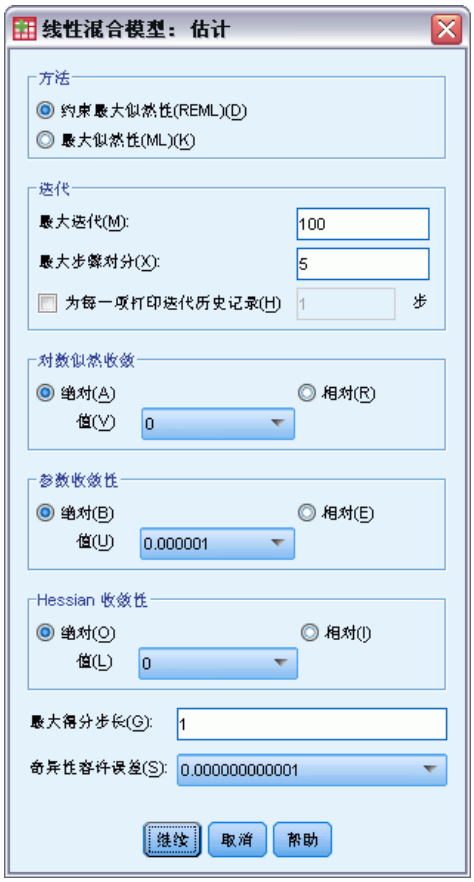
随机效应。没有缺省模型，因此必须显式指定随机效应。还可以构建嵌套或非嵌套项。也可以选择在随机效应模型中包括截距项。

可以指定多个随机效应模型。构建完第一个模型后，可以单击**下一个**构建下一个模型。单击**上一个**可以回滚现有模型。每个随机效应模型都被认为独立于所有其他随机效应模型；也就是说，会为每个随机效应模型计算一个单独的协方差矩阵。在同一随机效应模型中指定的项可以是相关的。

主体组。列出的变量就是您在“选择主体/重复变量”对话框中选择用作主体变量的变量。请选择这些变量的部分或全部来定义随机效应模型的主体。

线性混合模型：估计

图片 5-5
“线性混合模型：估计”对话框



方法。选择最大似然性或约束最大似然性。

迭代：

- **最大迭代次数。**指定一个非负整数。
- **最大步长对分。**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。
- **每隔 n 步打印迭代历史记录。**显示一个表，其中包含对数似然估计函数值和从第 0 次迭代（初始估计）开始每 n 次迭代的参数估计值。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。

对数似然性收敛性。如果对数似然函数的绝对变化或相对变化小于指定的非负值，则假定收敛。如果指定的值为 0，则不使用该标准。

参数收敛性。如果参数估计值的最大绝对变化或最大相对变化小于指定的非负值，则假定收敛。如果指定的值为 0，则不使用该标准。

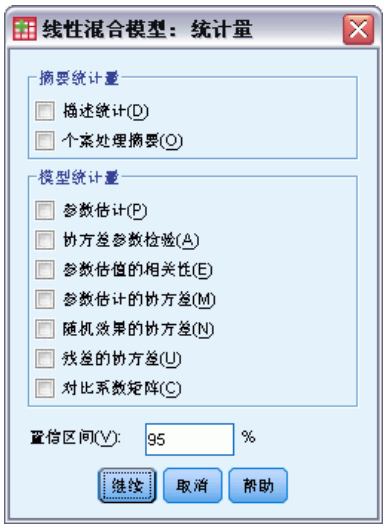
Hessian 收敛性。对于绝对指定，如果基于 Hessian 的统计量小于指定的值，则假定收敛。对于相对指定，如果统计量小于指定值与对数似然估计的绝对值的乘积，则假定收敛。如果指定的值为 0，则不使用该标准。

最大得分步长。请求使用 Fisher 评分算法达到迭代次数 n。指定一个正整数。

奇异性容许误差。这是在检查奇异性时用作容差的值。指定一个正值。

线性混合模型：统计

图片 5-6
“线性混合模型：统计量”对话框



摘要统计量。为以下项目生成表：

- **描述统计。**显示因变量和协变量（如果已指定）的样本大小、均值和标准差。将为各个不同的因子水平组合显示这些统计量。
- **个案处理摘要。**显示因子、重复测量变量、重复度量主体以及随机效应主体及其频率的排序值。

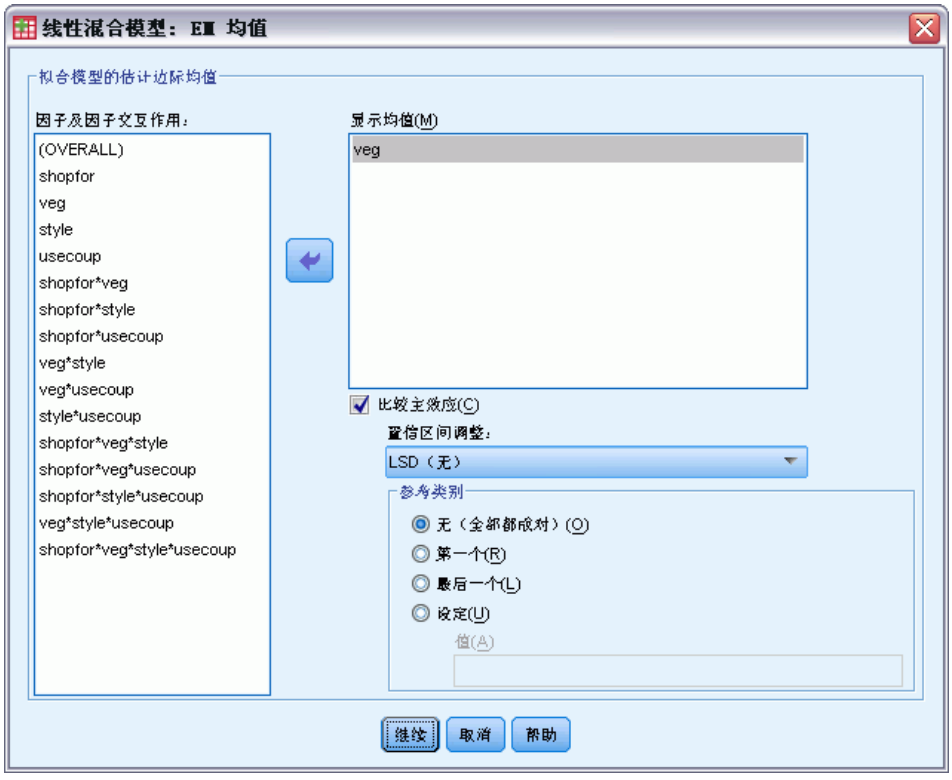
模型统计量。为以下项目生成表：

- **参数估计。**显示固定效应和随机效应参数估计值及其近似标准误。
- **协方差参数检验。**显示协方差参数的渐近标准误和 Wald 检验。
- **参数估值的相关性。**显示固定效应参数估计值的渐近相关矩阵。
- **参数估值协方差。**显示固定效应参数估计值的渐近协方差矩阵。
- **随机效果的协方差。**显示随机效应的估计的协方差矩阵。此选项仅当指定了至少一个随机效应时才可用。如果为随机效应指定了主体变量，则显示通用区组。
- **残差的协方差。**显示估计的残差协方差矩阵。此选项仅当指定了重复变量时才可用。如果指定了主体变量，则显示通用区组。
- **对比系数矩阵。**此选项显示用于检验固定效应和定制假设的可估计函数。

置信区间。只要构造了置信区间，就要使用此值。指定一个大于等于 0 且小于 100 的值。缺省值为 95。

线性混合模型：EM 均值

图片 5-7
“线性混合模型：EM 值”对话框



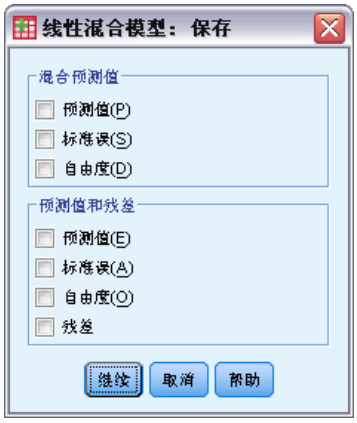
拟合模型的估计边际均值。此组允许您请求单元中因变量的模型预测的估计边际均值及其对于指定因子的标准误。此外，您还可以请求比较主效应的因子水平。

- **因子与因子交互。**此列表包含“固定”对话框中指定的因子与因子交互作用，以及一个 OVERALL 项。此列表中不包含根据协变量构建的模型项。

- **显示均值。**该过程将计算选入此列表的因子与因子交互作用的估计边际均值。如果选择了 OVERALL，则会显示因变量的估计边际均值，并会折叠所有因子。注意，选择的任何因子或因子交互作用都会一直选中，除非在主对话框中的“因子”列表中移去了关联变量。
- **比较主效应。**此选项允许您请求所选主效应水平的成对比较。“置信区间调整”允许调整置信区间和显著性值以便进行多重比较。可用的方法为 LSD（无调整）、Bonferroni 和 Sidak。最后，对于每个因子，您都可以选择进行比较的参考类别。如果不选择参考类别，则将构造所有成对比较。参考类别的选项可以首先指定、最后指定，也可以定制（在这种情况下，您需要输入参考类别的值）。

线性混合模型：保存

图片 5-8
“线性混合模型：保存”对话框



此对话框允许将各种模型结果保存到工作文件中。

混合预测值。保存与不具有这些效应的回归均值相关的变量。

- **预测值。**不具有随机效应的回归均值。
- **标准误。**估计值的标准误。
- **自由度。**与估计值相关联的自由度。

预测值和残差。保存与模型拟合值相关的变量。

- **预测值。**模型拟合值。
- **标准误。**估计值的标准误。
- **自由度。**与估计值相关联的自由度。
- **残差。**数据值减去预测值。

MIXED 命令的附加功能

使用命令语法语言还可以：

- 指定效应对比效应的线性组合或一个值的检验（使用 TEST 子命令）。
- 包括用户缺失值（使用 MISSING 子命令）。

- 计算协变量的指定值的估计边际均值（使用 EMMEANS 子命令的 WITH 关键字）。
- 比较交互的简单主效应（使用 EMMEANS 子命令）。

请参见命令语法参考以获取完整的语法信息。

章
6

广义线性模型

广义线性模型对一般线性模型进行了扩展，这样因变量通过指定的关联函数与因子和协变量线性相关。另外，该模型允许因变量呈非正态分布。它涵盖广泛使用的统计模型，例如用于正态分布响应的线性回归、用于二分类数据的 Logistic 模型、用于计数数据的对数线性模型、用于间隔检查生存数据的互补双对数模型，以及许多其他通过其非常通用的模型规划的统计模型。

示例。 运输公司可以使用广义线性模型，对在不同期间建造的一些轮船类型的损坏统计采用泊松回归，其结果模型可帮助确定哪些轮船类型最容易损坏。

汽车保险公司可以使用广义线性模型，对汽车损坏理赔采用 gamma 回归，其结果模型可帮助确定对理赔额度贡献最大的因素。

医疗研究人员可以使用广义线性模型，对间隔检查生存数据采用互补双对数回归，以预测医疗条件再次出现的时间。

数据。响应可以是尺度数据、计数数据、二分类数据或试验事件数据。假设因子是分类型的。假设协变量、尺度权重和偏移量是尺度型的。

假设。假设个案为独立观察值。

获取广义线性模型

从菜单中选择：

分析 > 广义线性模型 > 广义线性模型...

图片 6-1
“广义线性模型： 模型类型”选项卡

- 指定分布和关联函数（请参见下文获得有关各种选项的详细信息）。
- 在响应选项卡上选择一个因变量。
- 在预测选项卡上，选择用于预测因变量的因子和协变量。
- 在模型选项卡上，使用所选因子和协变量指定模型效应。

“模型类型”选项卡允许您为模型指定分布和关联函数，它为按响应类型分类的几种常用模型提供了快捷键。

模型类型

尺度响应。

- **线性。** 将正态指定为分布，将恒等指定为关联函数。
- **具有对数链接的 Gamma。** 将 Gamma 指定为分布，将对数指定为关联函数。

有序响应。

- **有序 logistic。**将多项（序数）指定为分布，将累积 logit 指定为关联函数。
- **有序 probit。**将多项（序数）指定为分布，将累积 probit 指定为关联函数。

计数。

- **泊松对数线性。** 将泊松指定为分布，将对数指定为关联函数。
- **具有对数链接的负二项式。** 将负二项式（拥有值为 1 的辅助参数）指定为分布，将对数指定为关联函数。要使过程估计辅助参数的值，指定一个拥有负二项式分布的定制模型，并在参数组中选择**估计值**。

二元响应或事件/试验数据。

- **二元 logistic。** 将二项式指定为分布，将 Logit 指定为关联函数。
- **二元 probit。** 将二项式指定为分布，将 Probit 指定为关联函数。
- **间隔检查生存。** 将二项式指定为分布，将互补双对数指定为关联函数。

混合。

- **具有对数链接的 Tweedie。** 将 Tweedie 指定为分布，将对数指定为关联函数。
- **具有恒等式链接的 Tweedie。** 将 Tweedie 指定为分布，将恒等指定为关联函数。

定制。指定您自己的分布和关联函数的组合。

分布

此选项指定因变量的分布。指定非正态分布和非恒等关联函数的功能是广义线性模型相对一般线性模型的重要改进。分布-关联函数可能存在多种组合，其中一些适合任何给定的数据集，因此可以根据先验理论的要求进行选择，或选择最合适的组合。

- **二项式。**此分布仅适合表示二元响应或事件数量的变量。
- **Gamma。**该分布适用于具有正尺度值并向更大的正值偏度的变量。如果数据值小于等于 0 或缺失，那么分析中不会使用相应的个案。
- **逆高斯。**该分布适用于具有正尺度值并向更大的正值偏度的变量。如果数据值小于等于 0 或缺失，那么分析中不会使用相应的个案。
- **负二项式。**该分布可以视为观察 k 成功所需的试验次数，适合具有非负整数值的变量。如果数据值是非整数、小于 0 或缺失，那么分析中不会使用相应的个案。负二项式分布辅助参数的固定值可以是大于等于 0 的任何值。您可将其设为固定值，或允许过程对其进行估计。辅助参数设置为 0 时，使用此分布相当于使用泊松分布。
- **正态。**该分布适合围绕某个中间值（均值）呈对称钟型分布的刻度变量。因变量必须是数值型变量。
- **泊松。**该分布可视为被观察事件在固定时间段内发生的次数，适合具有非负整数值的变量。如果数据值是非整数、小于 0 或缺失，那么分析中不会使用相应的个案。
- **Tweedie。** 此分布适合由 gamma 分布泊松混合表示的变量；之所以称为?混合?分布，是因为它兼具连续（取非负实数值）和离散分布（在单个值 0 处为正概率质量）的属性。因变量必须是数值型变量，数据值大于或等于零。如果数据值小于零或

缺失，那么分析中不会使用相应的个案。Tweedie 分布参数的固定值可以是任何大于 1 且小于 2 的数字。

- **多项式。**此分布适合表示序数响应的变量。因变量可以是数值或字符串，它必须至少有两个不同有效数据值。

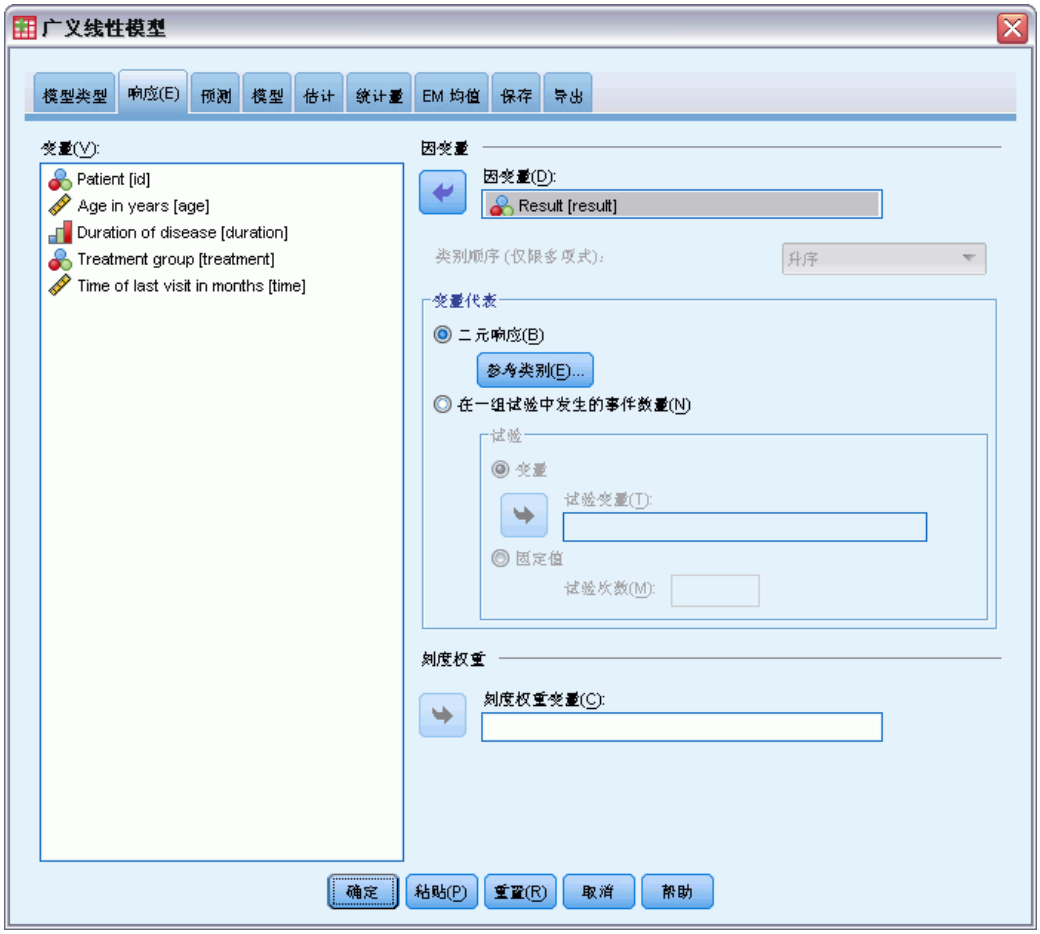
关联函数

联接函数是允许模型估计的因变量的转换。可用函数有：

- **恒等。**f(x)=x。因变量不转换。该关联可用于任何分布。
- **互补双对数。**f(x)=log(−log(1−x))。该函数只适用于二项式分布。
- **累积 Cauchit。**f(x) = tan(π (x - 0.5))，适用于每个响应类别的累积概率。该函数只适用于多项式分布。
- **累积互补双对数。**f(x)=ln(−ln(1−x))，适用于每个响应类别的累积概率。该函数只适用于多项式分布。
- **累积 logit。**f(x)=ln(x / (1−x))，适用于每个响应类别的累积概率。该函数只适用于多项式分布。
- **累积负双对数。**f(x)=−ln(−ln(x))，适用于每个响应类别的累积概率。该函数只适用于多项式分布。
- **累积 probit。**f(x)=Φ^{-1}(x)，适用于每个响应类别的累积概率，其中 Φ^{-1} 是逆标准正态累积分布函数。该函数只适用于多项式分布。
- **对数。**f(x)=log(x)。该关联可用于任何分布。
- **对数补数。**f(x)=log(1−x)。该函数只适用于二项式分布。
- **Logit。**f(x)=log(x / (1−x))。该函数只适用于二项式分布。
- **负二项式。**f(x)=log(x / (x+k^{-1}))，其中 k 是负二项式分布的辅助参数。该函数只适用于负二项式分布。
- **负双对数。**f(x)=−log(−log(x))。该函数只适用于二项式分布。
- **奇数幂。**f(x)=[(x/(1−x))$^{\alpha}$−1]/α，如果 α ≠ 0. f(x)=log(x)，如果 α=0. α 为必需的数字指定，且必须为实数。该函数只适用于二项式分布。
- **Probit。**f(x)=Φ^{-1}(x)，其中 Φ^{-1} 是逆标准正态累积分布函数。该函数只适用于二项式分布。
- **幂。**f(x)=x^{α}，如果 α ≠ 0. f(x)=log(x)，如果 α=0. α 为必需的数字指定，且必须为实数。该关联可用于任何分布。

广义线性模型响应

图片 6-2
“广义线性模型”对话框



在许多情况下，您可以只指定一个因变量；但是，仅采用两个值的变量和记录试验中事件的响应需要额外注意。

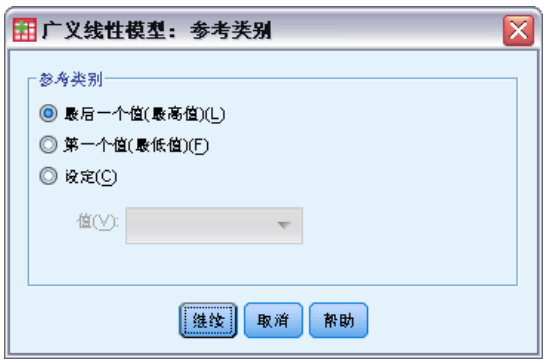
- **二元响应。**如果因变量仅采用两个值，可以为参数估计指定参考类别。二元响应变量可以是字符串或数值。
- **一组试验中发生的事件数量。**当响应是一系列试验中发生的事件数时，因变量包含事件数，您可以额外选择一个包含试验数的变量。或者，如果试验数在所有主体中都相同，则可以使用固定值指定试验。试验数应大于等于每个个案的事件数。事件应为非负整数，试验应为正整数。

对于有序多项模型，可以指定响应的类别顺序：升序、降序或数据（数据顺序意味着在数据中遇到的第一个值定义第一个类别，遇到的最后一个值定义最后一个类别。）

刻度权重。尺度参数是与响应方差相关的估计模型参数。尺度权重是“已知”值，可能因观察值的不同而异。如果指定了刻度权重变量，则对每个观察值，都会用与响应方差相关的尺度参数除以该尺度权重变量。分析中不使用尺度权重值小于等于 0 或缺失的个案。

广义线性模型：参考类别

图片 6-3
“广义线性模型：参考类别”对话框

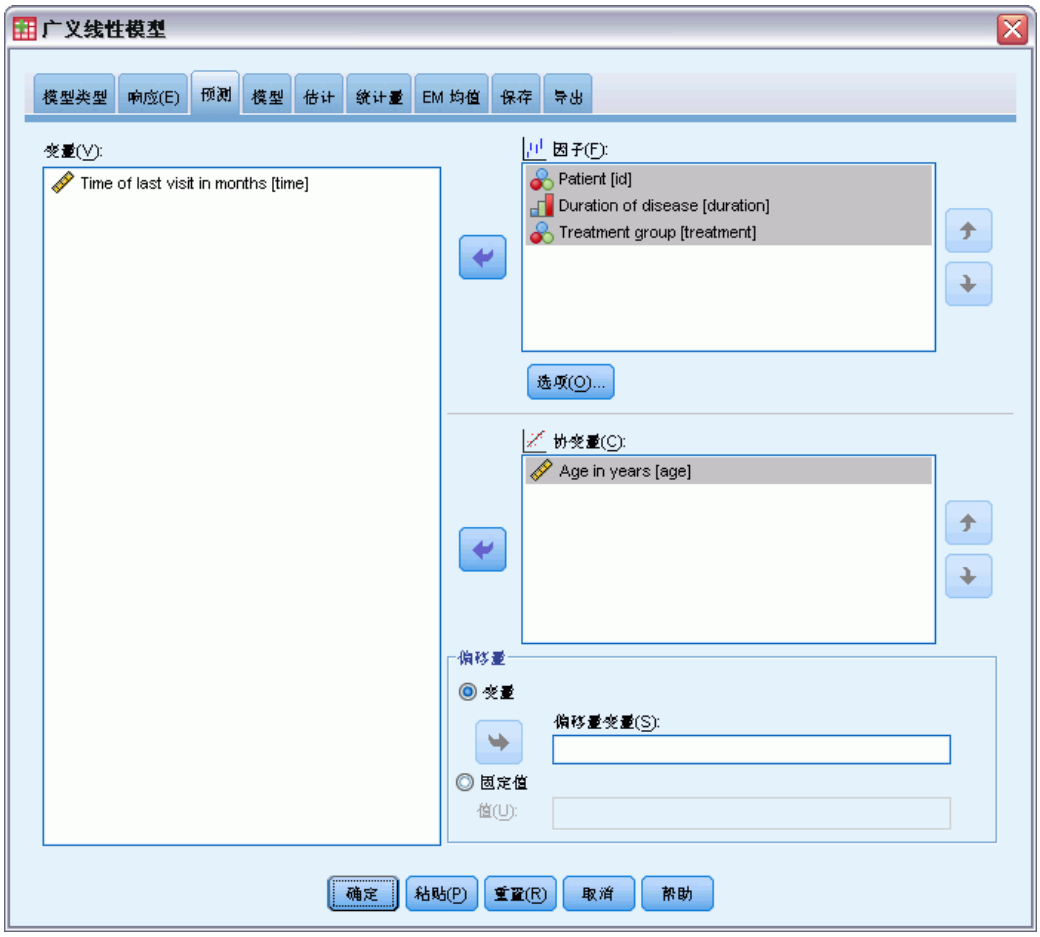


对于二元响应，可以为因变量选择参考类别。这会影响某些输出，如参数估计值和保存的值，但不应更改模型拟合。例如，如果您的二元响应取值 0 和 1：

- 缺省情况下，此过程将最后一个值（最高值）或 1 作为参考类别。在这种情况下，模型保存的概率估计给定个案取值 0 的概率，参数估计值应解释为与类别 0 的似然估计相关。
- 如果您指定第一个值（最低值）或 0 作为参考类别，则模型保存的概率估计给定个案取值 1 的概率。
- 如果您指定定制值并且变量已定义了标签，则可以通过从列表中选择值来设置参考类别。在指定模型的过程中并不确定某一特定变量的编码方式时，这种方法非常方便。

广义线性模型：预测变量

图片 6-4
Generalized Linear Models:“预测变量”选项卡



在“预测”选项卡中可以指定用于生成模型效应的因子和协变量并指定可选偏移。

因子。因子是分类预测变量，可以是数值或字符串。

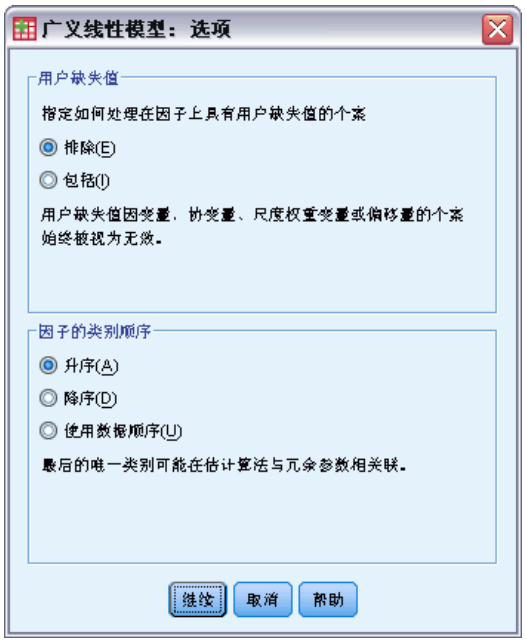
协变量。协变量为刻度预测变量，必须为数值。

注意：当响应为二元格式的二项式时，该过程通过子体来计算偏差和卡方拟合优度统计量，这些子体基于对所选因子和协变量观察值的交叉分类。您应在过程的多次运行中保持相同的预测变量集，以确保一致的子体数量。

偏移量。偏移项是“结构化”预测变量。模型不估计该预测变量的系数，但假定其值为 1；因此，偏移值只是简单地加到因变量的线性预测变量中。这在每个个案对于被观察事件都可能具有不同显现水平的泊松回归模型中尤其有用。例如，当对各个驾驶员事故率进行建模时，3 年驾驶经历出现 1 次事故和 25 年出现 1 次事故的驾驶员之间有着重大的差别！如果将驾驶员经历纳入偏移项，则事故数可以建模为泊松响应。

广义线性模型：选项

图片 6-5
“广义线性模型:选项”对话框



这些选项适用于“预测”选项卡上指定的所有因子。

用户缺失值。要在分析中包含个案，因子必须具有有效值。通过这些控制可以决定是否将用户缺失值在因子变量中视为有效值。

类别顺序。该选项与确定因子的最终水平有关，该水平可能与估计算法中的冗余参数相关。更改类别顺序可更改因子水平效应的值，因为这些参数估计值相对于“最终”水平进行计算。因子可以按从最低值到最高值的升序排序，按从最高值到最低值的降序排序，或者按“数据顺序”排序。这意味着在数据中遇到的第一个值定义第一个类别，遇到的最后一个唯一值定义最后一个类别。

广义线性模型：模型

图片 6-6
Generalized Linear Models:“模型”选项卡

指定模型效应。缺省模型是仅截距模型，因此必须明确指定其他模型效应。还可以构建嵌套或非嵌套项。

非嵌套项

对于选定因子和协变量：

主效应。为每个选定的变量创建主效应项。

交互。为所有选定变量创建最高级交互项。

因子。创建选定变量的所有可能的交互和主效应。

所有二阶。创建选定变量的所有可能的二阶交互。

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。由于每位顾客只频繁光顾某一位置的商店，因此 Customer 效应可以说是**嵌套在** Store location 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。

多项有序分布的模型没有单个截距项；而包含定义相邻类别之间转换点的阈值参数。模型中通常包含有阈值。

广义线性模型：估计

图片 6-7
Generalized Linear Models:“估计”选项卡

参数估计。该组中的控件允许指定估计方法并提供参数估计值的初始值。

- **方法。**可以选择参数估计方法。可以选择的方法包括 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，则算法将继续执行 Newton-Raphson 方法。
- **刻度参数方法。**可以选择尺度参数估计方法。最大似然法可联合估计尺度参数和模型效应；请注意，如果响应具有负二项式、泊松、二项式或多项式分布，则此选项无效。偏差和 Pearson 卡方选项从这些统计量的值估计尺度参数。或者，可以为尺度参数指定固定值。
- **初始值。**该过程将自动计算参数的初始值。也可以指定参数估计值的初始值。
- **协方差矩阵。**基于模型的估计是 Hessian 矩阵的广义逆负矩阵。健壮性（也称为 Huber/White/sandwich）估计是“改正”的基于模型的估计，即使错误地指定了方差和关联函数，也能提供对协方差的一致估计。

迭代。

- **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。
- **最大步骤对分。**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。
- **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。此选项可用于二元格式的多项式响应和二项式响应。

收敛性准则。

- **参数收敛。**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。
- **对数似然估计收敛。**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。
- **Hessian 收敛性。**对于“绝对值”指定，如果基于 Hessian 收敛的统计量小于指定的正值，则假定收敛。对于“相对”指定，如果统计量小于指定的正值和对数似然估计的绝对值的乘积，则假定收敛。

奇异性容许误差。奇异（非可逆）矩阵具有线性相关列，对估计算法可能产生严重问题。即使近似奇异的矩阵也可导致不良结果，因此该过程会将行列式小于容许误差的矩阵作为奇异矩阵对待。指定一个正值。

广义线性模型：初始值

图片 6-8
“广义线性模型：初始值”对话框

如果指定了初始值，则必须为模型中的所有参数（包括冗余参数）提供初始值。在数据集中，变量从左到右的顺序必须为：RowType_、VarName_、P1、P2、…，其中 RowType_ 和 VarName_ 是字符串变量，P1、P2、… 是对应于已排序参数列表的数值变量。

- 初始值在变量 RowType_ 具有值 EST 的记录上提供；实际初始值在变量 P1、P2、… 下提供。该过程忽略 RowType_ 具有非 EST 值的所有记录以及 RowType_ 的第一次等于 EST 以外的所有记录。
- 截距（如果模型中包含）或阈值参数（如果响应具有多项式分布）必须是列出的第一个初始值。

- 尺度参数和负二项式参数（如果响应具有负二项式分布）必须是指定的最后一个初始值。
- 如果“拆分文件”有效，则变量必须以拆分文件变量开始（以创建“拆分文件”时指定的顺序），后跟 RowType_、VarName_、P1、P2、…，如上所述。拆分在指定数据集中的发生顺序必须与原始数据集中的顺序相同。

注意：变量名 P1、P2、… 不是必需的；过程接受参数的任何有效变量名，因为变量与参数的映射基于变量位置，而不是变量名。超过最后一个参数的所有变量都会被忽略。

初始值的文件结构与将模型导出为数据时使用的结构相同；因此，可以使用一次过程运行的最终值作为下一次运行的输入。

广义线性模型：统计量

图片 6-9
Generalized Linear Models:“统计量”选项卡

模型作用。

- **分析类型。**指定要生成的分析类型。类型 I 分析通常适合于对模型中预测变量的排序具有先验理由的情况，而类型 III 应用范围更广。根据在“卡方统计量”组中的选择，计算 Wald 或似然比统计。

- **置信区间。**指定一个大于 50 且小于 100 的置信度。Wald 区间基于参数呈渐近正态分布的假设之上；截面似然区间更加准确但需要进行大量的计算。截面似然区间的容差水平可作为标准，用以停止在区间计算所采用的迭代算法。
- **对数似然函数。**它控制对数似然函数的显示格式。整个函数包含一个相对于参数估计恒定的额外项；它对参数估计没有任何影响，并且在某些软件产品中不会显示。

打印。可用输出如下：

- **个案处理摘要。**显示分析和“相关数据摘要”表中包含的个案和从中排除的个案的数量和百分比。
- **描述统计。**显示有关因变量、协变量和因子的描述统计和摘要信息。
- **模型信息。**显示数据文件名称、因变量或事件和试验变量、偏移变量、尺度权重变量、概率分布和关联函数。
- **拟合度统计。**显示离差和尺度化离差、Pearson 卡方和尺度化 Pearson 卡方、对数似然估计、AIC 准则 (AIC)、有限样本校正 AIC (AICC)、BIC 准则 (BIC) 以及 CAIC 准则 (CAIC)。
- **模型摘要统计。**显示模型拟合检验，包括用于模型拟合 Omnibus 检验的似然比统计量，以及每种效应的类型 I 或 III 对比的统计量。
- **参数估计。**显示参数估计值和相应的检验统计和置信区间。除了显示原始参数估计值之外，还可以选择显示取幂参数估计值。
- **参数估计的协方差矩阵。**显示估计参数协方差矩阵。
- **参数估计的相关性矩阵。**显示估计参数相关矩阵。
- **对比系数 (L) 矩阵。**为缺省效应和估计边际均值显示对比系数（如果“EM 均值”选项卡上要求）。
- **常规可估计函数。**显示生成对比系数 (L) 矩阵的矩阵。
- **迭代历史记录。**显示参数估计值和对数似然估计的迭代历史记录，并打印对梯度矢量和 Hessian 矩阵的最后一次评估。迭代历史记录表从第 0 次迭代（初始估计值）开始为每 n 次迭代显示参数估计值，其中，n 代表打印区间的值。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。
- **Lagrange 乘数检验。**显示 Lagrange 乘数检验统计量，用于为正态、gamma、逆高斯和 Tweedie 分布评估使用离差或 Pearson 卡方计算或者设置为固定值的尺度参数的有效性。对于负二项式分布，它检验固定辅助参数。

广义线性模型：EM 均值

图片 6-10
Generalized Linear Models: “EM 均值”选项卡

此选项卡用于显示因子水平和因子交互的估计边际均值。您还可以要求显示整体估计均值。估计边际均值对有序多项模型不可用。

因子和交互。此列表包含“预测变量”选项卡上指定的因子和“模型”选项卡上指定的因子交互。此列表中不包含协变量。可以直接从此列表选择项，或者使用依据 * 按钮将项组合到交互项中。

显示以下项的均值。计算所选因子和因子交互的估计均值。对比决定如何设定假设检验以比较估计均值。简单对比需要一个用于比较其他项的参考类别或因子水平。

- **配对方式。**为指定或默示的因子的所有水平组合计算成对比较。这是因子交互的唯一可用对比方法。
- **简单散点图.** 将每个水平的均值与指定水平的均值进行比较。当存在控制组时，此类对比很有用。
- **偏差。**因子的每个水平与总均值比较。偏移对比不是正交的。

- **差分**. 将每个水平（第一个除外）的均值与先前水平的均值进行比较。有时候将其称为逆 Helmert 对比。
- **Helmert**. 将因子的每个水平的均值（最后一个水平除外）与后续水平的均值进行比较。
- **重复**. 将每个水平的均值（最后一个水平除外）与后续水平的均值进行比较。
- **多项式**. 比较线性作用、二次作用、三次作用等等。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。

标度。对于响应，可根据因变量的原始刻度计算估算边际均值；对于线性预测变量，根据关联函数变换的因变量计算估计边际均值。

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。
- **Bonferroni**. 此方法针对检验多个对比这一事实调整观测的显著性水平。
- **连续 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 方法更严密的界限。
- **连续 Sidak**. 这是按顺序逐步降低的拒绝 Sidak 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。

广义线性模型：保存

图片 6-11
Generalized Linear Models:“保存”选项卡

选中的项以指定的名称保存；可以选择用新变量覆盖现有的同名变量，或为新变量名添加后缀使其成为唯一名称以避免名称冲突。

- **响应均值的预测值。**以初始响应度规保存每个个案的模型预测值。当响应分布是二项式，并且因变量为二元时，过程会保存预测概率。当响应分布是多项式时，项目标签变成**累积预测概率**，并且过程保存响应每个类别的累积预测概率（最后一个除外），直至达到要保存的指定类别数量。
- **响应均值的置信区间下限**保存响应均值置信区间的下限。当响应分布是多项式时，项目标签变成**累积预测概率的置信区间下限**，并且过程保存响应每个类别的下限（最后一个除外），直至达到要保存的指定类别数量。
- **响应均值的置信区间上限。**保存响应均值置信区间的上限。当响应分布是多项式时，项目标签变成**累积预测概率的置信区间上限**，并且过程保存响应每个类别的上限（最后一个除外），直至达到要保存的指定类别数量。
- **预测类别。**对于具有二项式分布和二元因变量或多项式分布的模型，该过程保存每个个案的预测响应类别。此选项不支持其他响应分布。

- **线性预测器的预测值。**以线性预测变量的度规保存每个个案的模型预测值（通过指定的关联函数转换的响应）。当响应分布是多项式时，过程保存响应每个类别的预测值（最后一个除外），直至达到要保存的指定类别数量。
- **线性预测变量的预测值的估计标准误。**当响应分布是多项式时，过程保存响应每个类别的估计标准误（最后一个除外），直至达到要保存的指定类别数量。

如果响应分布是多项式，则以下各项不可用。

- **Cook 距离.** 在特定个案从回归系数的计算中排除的情况下，所有个案的残差变化幅度的测量。较大的 Cook 距离表明从回归统计量的计算中排除个案之后，系数会发生根本变化。
- **杠杆值.** 度量某个点对回归拟合的影响。集中的杠杆值范围为从 0（对拟合无影响）到 (N-1)/N。
- **原始残差.** 观察值与模型预测值之间的差。
- **Pearson 残差。**个案对 Pearson 卡方统计量的贡献的平方根（带原始残差的符号）。
- **标准化 Pearson 残差。**尺度参数与 1–（个案的杠杆）的乘积的逆的平方根乘以 Pearson 残差。
- **偏差残差。**个案对偏差统计量的贡献的平方根（带原始残差的符号）。
- **标准化偏差残差。**尺度参数与 1–（个案的杠杆）的乘积的逆的平方根乘以偏差残差。
- **似然残差。**标准化 Pearson 和标准化偏差残差的平方的加权平均值（基于个案杠杆）的平方根，带原始残差的符号。

广义线性模型：导出

图片 6-12
Generalized Linear Models:“导出”选项卡

将模型导出为数据。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、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。

- **拆分变量。**定义拆分的任何变量（如果使用）。
- **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）、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单独个案。
- **VarName_。** 对于行类型 COV 或 CORR，取值为 P1、P2、...，对应于所有估计模型参数（除尺度或负二项式参数外）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（必要时可包括尺度或负二项式参数）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、显著性水平和残差自由度设为系统缺失值。

对于尺度参数，协方差、相关性、显著性水平和自由度均设为系统缺失值。如果尺度参数通过最大似然进行估计，则给出标准误；否则，它被设为系统缺失值。

对于负二项式参数，协方差、相关性、显著性水平和自由度均设为系统缺失值。如果负二项式参数通过最大似然进行估计，则给出标准误；否则，它被设为系统缺失值。

如果存在拆分，则参数列表必须在所有拆分之间进行累积。在给定拆分中，某些参数可能是无关的；但这不同于冗余。对于无关参数，所有协方差或相关性、参数估计值、标准误、显著性水平和自由度均设为系统缺失值。

可以使用此矩阵文件作为进一步进行模型估计的初始值；注意，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。尽管如此，您必须注意，在该矩阵文件中的所有参数对于读取此文件的过程具有相同的含义。

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。

GENLIN 命令的附加功能

使用命令语法语言还可以：

- 将参数估计值的初始值指定为数字列表（使用 CRITERIA 子命令）。
- 在计算估计边际均值时修正值不同于其均值的协变量（使用 EMMEANS 子命令）。
- 指定估计边际均值的定制多项式对比（使用 EMMEANS 子命令）。
- 指定因子的一个子集，以显示该子集的估计边际均值，并使用指定的对比类型对其进行比较（使用 EMMEANS 子命令的 TABLES 和 COMPARE 关键字）。

请参阅命令语法参考以获取完整的语法信息。

章
7

广义估计方程

广义估计方程过程对广义线性模型进行了扩展，以允许分析重复的测量或其他相关观察数据，例如聚类数据。

示例。公共卫生官员可以使用广义估计方程，在空气污染对儿童影响研究中采用重复度量 Logistic 回归模型。

数据。响应可以是尺度数据、计数数据、二分类数据或试验事件数据。假设因子是分类型的。假设协变量、尺度权重和偏移量是尺度型的。用于定义主体或主体内重复度量的变量不能用于定义响应，但可以在模型中发挥其他作用。

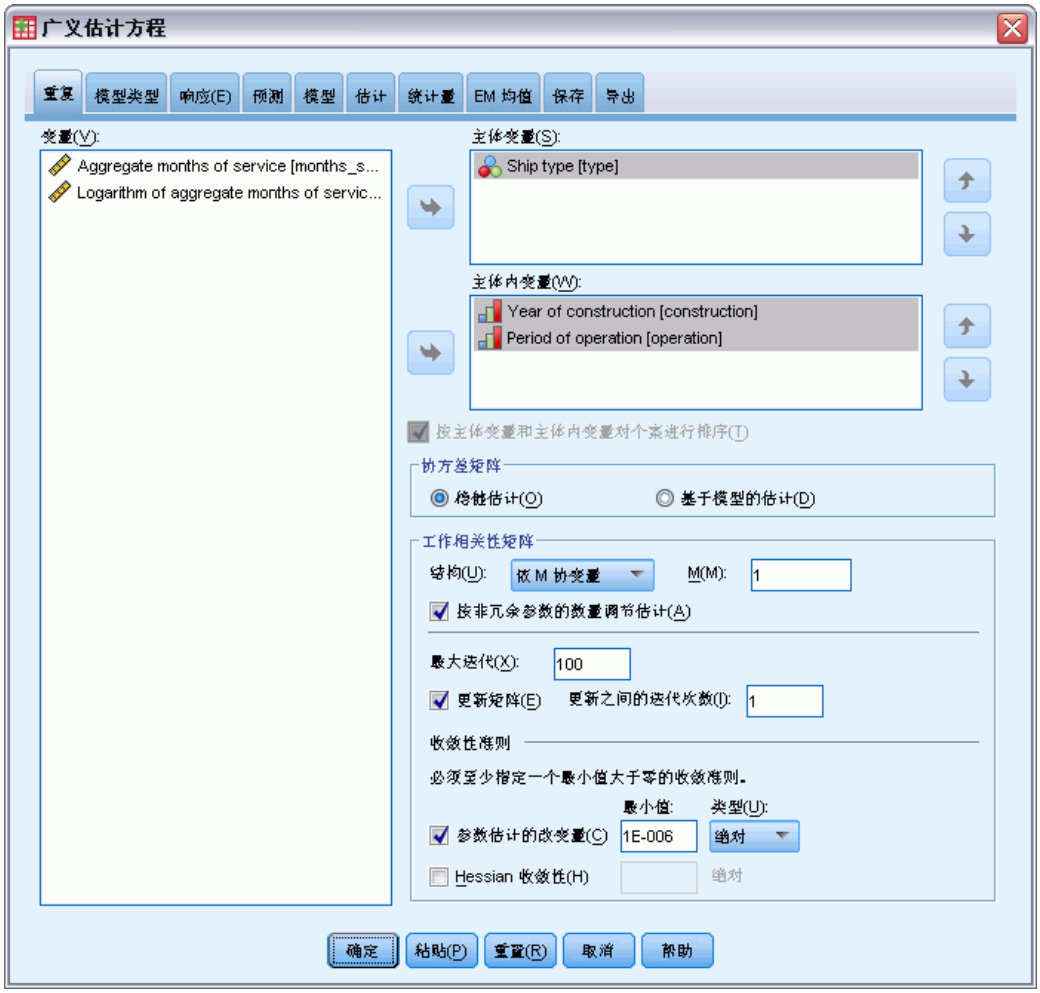
假设。假设各个个案在主体内部是相关的，在主体之间是独立的。表示主体内相关性的相关矩阵作为模型的一部分进行估计。

获得广义估计方程

从菜单中选择：

分析 > 广义线性模型 > 广义估计方程...

图片 7-1
广义估计方程：“重复”选项卡



- 选择一个或多个主体变量（参见下文了解更多选项）。

  指定变量的值组合应唯一定义数据集中的**主体**。例如，单个病人 ID 变量应足以定义一个医院内的主体，但如果病人标识号在医院间不唯一，则需要使用医院 ID 和病人 ID 的组合。在重复度量设置中，将为每个主体记录多个观察数据，因此每个主体可能在数据集中占用多个个案。

- 在模型类型选项卡上，指定分布和关联函数。
- 在响应选项卡上选择一个因变量。
- 在预测选项卡上，选择用于预测因变量的因子和协变量。
- 在模型选项卡上，使用所选因子和协变量指定模型效应。

或者，可以在“重复”选项卡上指定：

主体内变量。主体内变量值的组合定义主体中度量的顺序；因此，主体内变量和主体变量的组合唯一定义每个度量。例如，时间、医院 ID 和病人 ID 的组合为每个个案定义特定医院中特定病人的一次就诊。

如果数据集已经排序，每个主体的重复度量因而按正确顺序在连续个案段中发生，则并不严格要求必须指定主体内变量，并且您可以取消选择**按个体变量和主体内变量对个案进行排序**并保存执行（临时）排序所需的处理时间。通常，利用主体内变量确保度量的正确顺序是很好的方法。

主体变量和主体内变量不能用于定义响应，但它们可以在模型中执行其他功能。例如，医院 ID 可用作模型中的因子。

协方差矩阵。基于模型的估计是 Hessian 矩阵的广义逆负矩阵。健壮性估计（也称为 Huber/White/sandwich 估计）是“改正”的基于模型的估计，即使错误地指定了工作相关矩阵，也能提供对协方差的一致估计。该规范适用于广义估计方程的线性模型部分中的参数，而估计选项卡上的规范只适用于初始广义线性模型。

工作相关性矩阵。此相关矩阵表示主体内相关性。其大小由度量数决定，因此也由主体内变量的值组合决定。您可以指定以下结构之一：

- **独立。**重复度量不相关。
- **AR(1)。**重复度量具有一阶自回归关系。任意两个元素之间的相关性对于相邻元素为 ρ，对于由第三个元素分隔的元素为 ρ^2，依此类推。ρ 受到约束，以使 $-1<\rho<1$。
- **可交换。**此结构在元素之间具有同质相关性。又称为复合对称结构。
- **依 M 协变量。**连续的测量具有共同的相关系数，由第三个度量分隔的测量对具有共同的相关系数，依此类推，直到由 m–1 个其他度量分隔的测量对。具有更多分隔的测量假设为不相关。选择此结构时，请指定小于工作相关矩阵的阶的 m 值。
- **未结构化。**这是一个非常一般的相关矩阵。

缺省情况下，过程将根据非冗余参数的数目调节相关估计值。如果希望估计值不会随着数据中的主体级重复变化而发生变化，则可能需要去掉此调节功能。

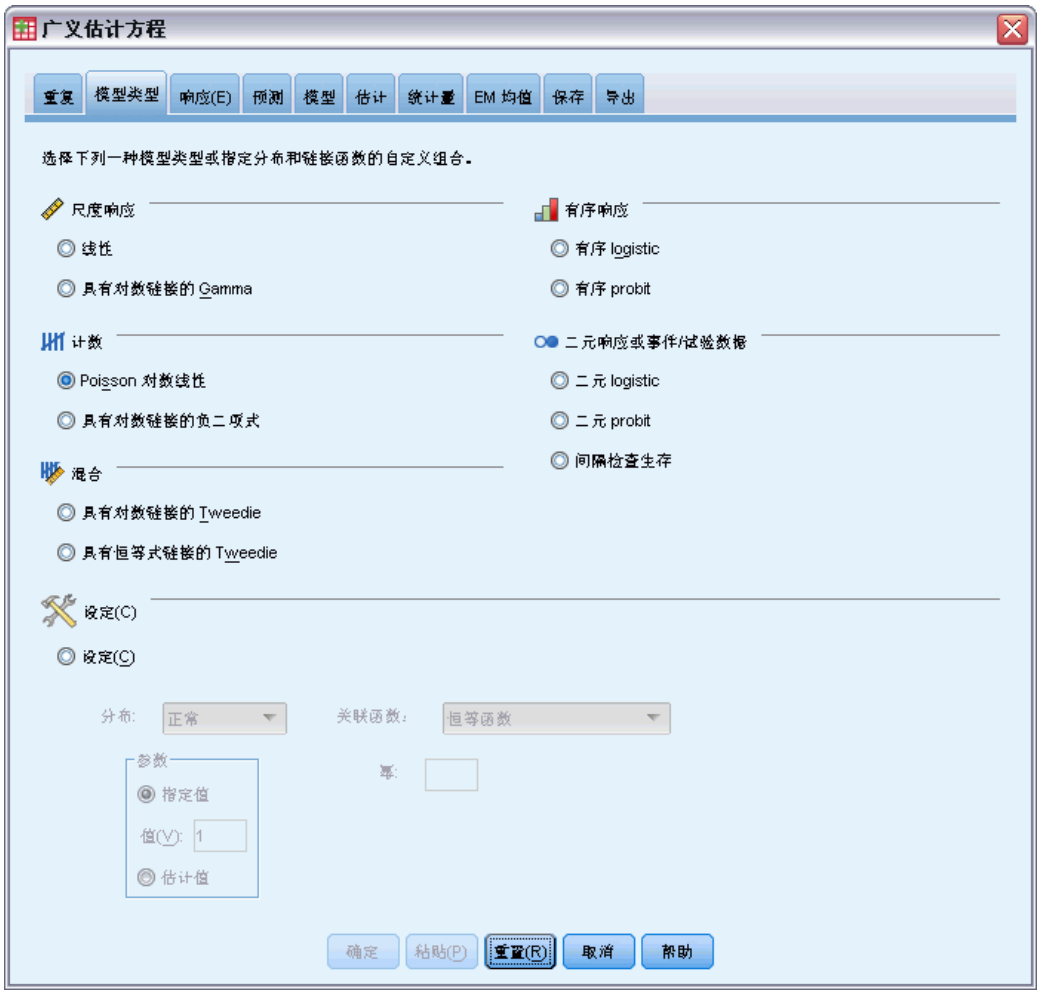
- **最大迭代次数。**广义估计方程算法将执行的最大迭代次数。指定一个非负整数。该规范适用于广义估计方程的线性模型部分中的参数，而估计选项卡上的规范只适用于初始广义线性模型。
- **更新矩阵。**工作相关矩阵中的元素将根据参数估计值进行估计，参数估计值在算法的每次迭代中更新。如果工作相关矩阵完全没有更新，则在整个估计过程中将使用初始工作相关矩阵。如果该矩阵进行了更新，则可以指定更新工作相关矩阵元素的迭代间隔。指定大于 1 的值可缩短处理时间。

收敛性准则。这些规范适用于广义估计方程的线性模型部分中的参数，而估计选项卡上的规范只适用于初始广义线性模型。

- **参数收敛。**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。
- **Hessian 收敛性。**如果基于 Hessian 的一个统计量小于指定值（必须为正值），则认为收敛。

广义估计方程：模型类型

图片 7-2
广义估计方程：“模型类型”选项卡



“模型类型”选项卡允许您为模型指定分布和关联函数，它为按响应类型分类的几种常用模型提供了快捷键。

模型类型

尺度响应。

- **线性。** 将正态指定为分布，将恒等指定为关联函数。
- **具有对数链接的 Gamma。** 将 Gamma 指定为分布，将对数指定为关联函数。

有序响应。

- **有序 logistic。**将多项（序数）指定为分布，将累积 logit 指定为关联函数。
- **有序 probit。**将多项（序数）指定为分布，将累积 probit 指定为关联函数。

计数。

- **泊松对数线性。** 将泊松指定为分布，将对数指定为关联函数。
- **具有对数链接的负二项式。** 将负二项式（拥有值为 1 的辅助参数）指定为分布，将对数指定为关联函数。要使过程估计辅助参数的值，指定一个拥有负二项式分布的定制模型，并在参数组中选择**估计值**。

二元响应或事件/试验数据。

- **二元 logistic。** 将二项式指定为分布，将 Logit 指定为关联函数。
- **二元 probit。** 将二项式指定为分布，将 Probit 指定为关联函数。
- **间隔检查生存。** 将二项式指定为分布，将互补双对数指定为关联函数。

混合。

- **具有对数链接的 Tweedie。** 将 Tweedie 指定为分布，将对数指定为关联函数。
- **具有恒等式链接的 Tweedie。** 将 Tweedie 指定为分布，将恒等指定为关联函数。

定制。指定您自己的分布和关联函数的组合。

分布

此选项指定因变量的分布。指定非正态分布和非恒等关联函数的功能是广义线性模型相对一般线性模型的重要改进。分布-关联函数可能存在多种组合，其中一些适合任何给定的数据集，因此可以根据先验理论的要求进行选择，或选择最合适的组合。

- **二项式。**此分布仅适合表示二元响应或事件数量的变量。
- **Gamma。**该分布适用于具有正尺度值并向更大的正值偏度的变量。如果数据值小于等于 0 或缺失，那么分析中不会使用相应的个案。
- **逆高斯。**该分布适用于具有正尺度值并向更大的正值偏度的变量。如果数据值小于等于 0 或缺失，那么分析中不会使用相应的个案。
- **负二项式。**该分布可以视为观察 k 成功所需的试验次数，适合具有非负整数值的变量。如果数据值是非整数、小于 0 或缺失，那么分析中不会使用相应的个案。负二项式分布辅助参数的固定值可以是大于等于 0 的任何值。您可将其设为固定值，或允许过程对其进行估计。辅助参数设置为 0 时，使用此分布相当于使用泊松分布。
- **正态。**该分布适合围绕某个中间值（均值）呈对称钟型分布的刻度变量。因变量必须是数值型变量。
- **泊松。**该分布可视为被观察事件在固定时间段内发生的次数，适合具有非负整数值的变量。如果数据值是非整数、小于 0 或缺失，那么分析中不会使用相应的个案。
- **Tweedie。** 此分布适合由 gamma 分布泊松混合表示的变量；之所以称为?混合?分布，是因为它兼具连续（取非负实数值）和离散分布（在单个值 0 处为正概率质量）的属性。因变量必须是数值型变量，数据值大于或等于零。如果数据值小于零或缺失，那么分析中不会使用相应的个案。Tweedie 分布参数的固定值可以是任何大于 1 且小于 2 的数字。
- **多项式。**此分布适合表示序数响应的变量。因变量可以是数值或字符串，它必须至少有两个不同有效数据值。

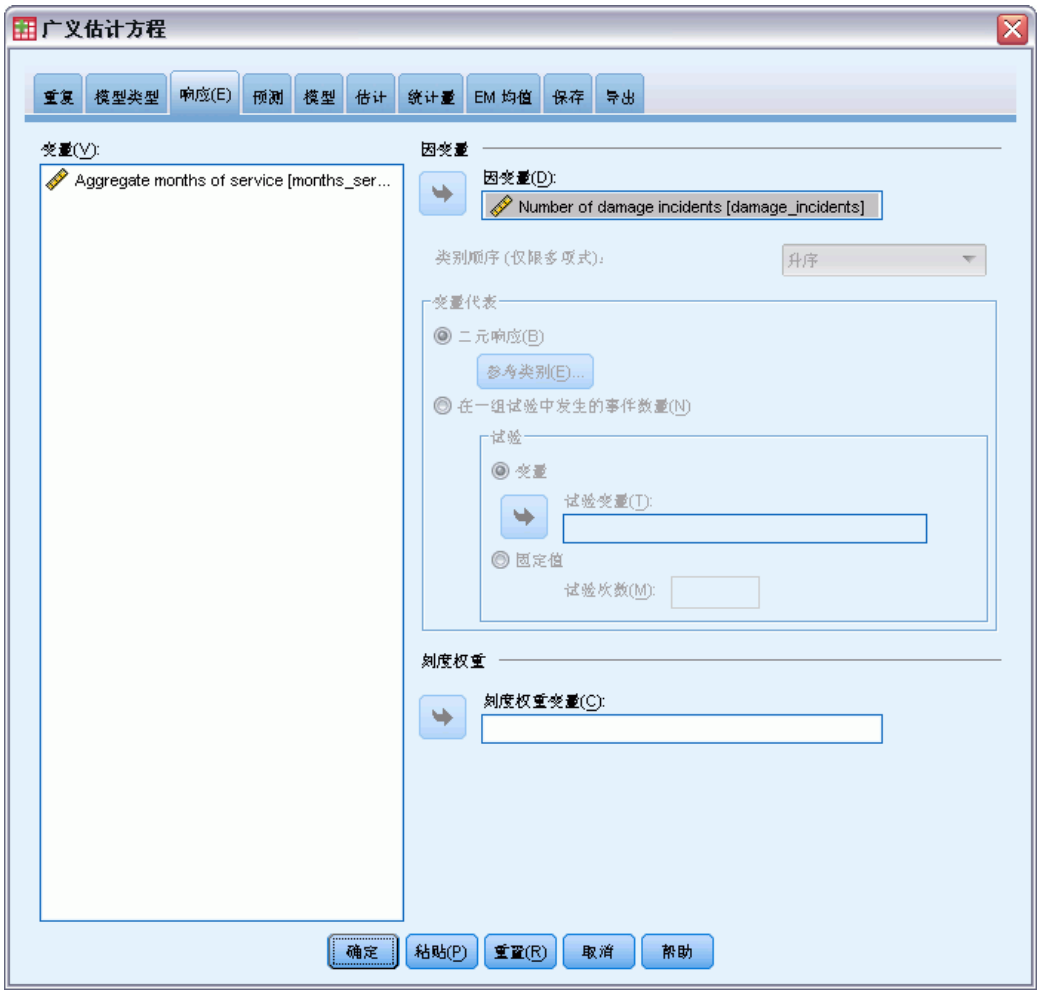
联接函数

联接函数是允许模型估计的因变量的转换。可用函数有：

- **恒等。**f(x)=x。因变量不转换。该关联可用于任何分布。
- **互补双对数。**f(x)=log(−log(1−x))。该函数只适用于二项式分布。
- **累积 Cauchit。**f(x) = tan(π (x - 0.5))，适用于每个响应类别的累积概率。该函数只适用于多项式分布。
- **累积互补双对数。**f(x)=ln(−ln(1−x))，适用于每个响应类别的累积概率。该函数只适用于多项式分布。
- **累积 logit。**f(x)=ln(x / (1−x))，适用于每个响应类别的累积概率。该函数只适用于多项式分布。
- **累积负双对数。**f(x)=−ln(−ln(x))，适用于每个响应类别的累积概率。该函数只适用于多项式分布。
- **累积 probit。**f(x)=Φ^{-1}(x)，适用于每个响应类别的累积概率，其中 Φ^{-1} 是逆标准正态累积分布函数。该函数只适用于多项式分布。
- **对数。**f(x)=log(x)。该关联可用于任何分布。
- **对数补数。**f(x)=log(1−x)。该函数只适用于二项式分布。
- **Logit。**f(x)=log(x / (1−x))。该函数只适用于二项式分布。
- **负二项式。**f(x)=log(x / (x+k^{-1}))，其中 k 是负二项式分布的辅助参数。该函数只适用于负二项式分布。
- **负双对数。**f(x)=−log(−log(x))。该函数只适用于二项式分布。
- **奇数幂。**f(x)=[(x/(1−x))$^{\alpha}$−1]/α，如果 α ≠ 0. f(x)=log(x)，如果 α=0. α 为必需的数字指定，且必须为实数。该函数只适用于二项式分布。
- **Probit。**f(x)=Φ^{-1}(x)，其中 Φ^{-1} 是逆标准正态累积分布函数。该函数只适用于二项式分布。
- **幂。**f(x)=x^{α}，如果 α ≠ 0. f(x)=log(x)，如果 α=0. α 为必需的数字指定，且必须为实数。该关联可用于任何分布。

广义估计方程：响应

图片 7-3
广义估计方程：“响应”选项卡



在许多情况下，您可以只指定一个因变量；但是，仅采用两个值的变量和记录试验中事件的响应需要额外注意。

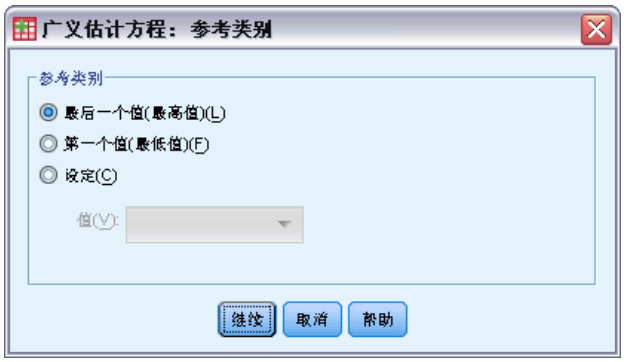
- **二元响应。**如果因变量仅采用两个值，可以为参数估计指定参考类别。二元响应变量可以是字符串或数值。
- **一组试验中发生的事件数量。**当响应是一系列试验中发生的事件数时，因变量包含事件数，您可以额外选择一个包含试验数的变量。或者，如果试验数在所有主体中都相同，则可以使用固定值指定试验。试验数应大于等于每个个案的事件数。事件应为非负整数，试验应为正整数。

对于有序多项模型，可以指定响应的类别顺序：升序、降序或数据（数据顺序意味着在数据中遇到的第一个值定义第一个类别，遇到的最后一个值定义最后一个类别。）

刻度权重。尺度参数是与响应方差相关的估计模型参数。尺度权重是“已知”值，可能因观察值的不同而异。如果指定了刻度权重变量，则对每个观察值，都会用与响应方差相关的尺度参数除以该尺度权重变量。分析中不使用尺度权重值小于等于 0 或缺失的个案。

广义估计方程：参考类别

图片 7-4
“广义估计方程：参考类别”对话框

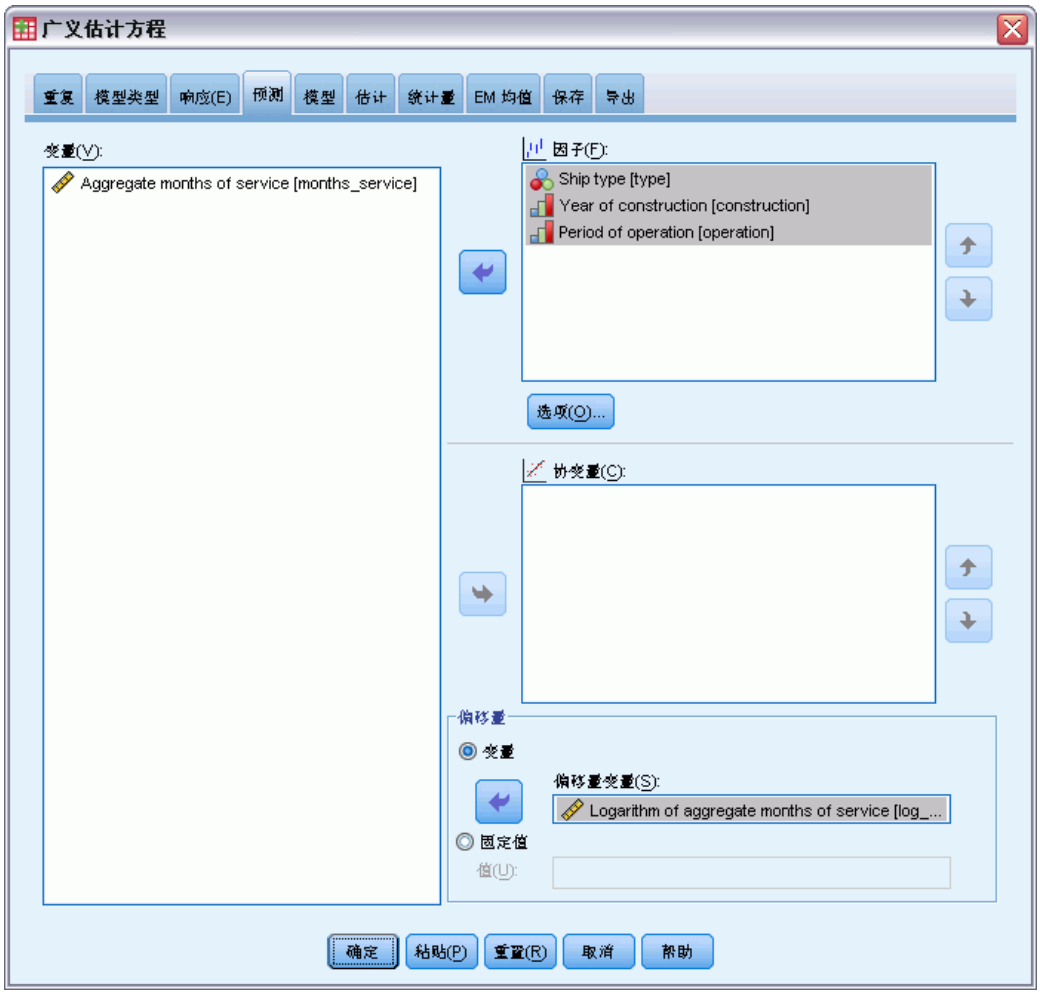


对于二元响应，可以为因变量选择参考类别。这会影响某些输出，如参数估计值和保存的值，但不应更改模型拟合。例如，如果您的二元响应取值 0 和 1：

- 缺省情况下，此过程将最后一个值（最高值）或 1 作为参考类别。在这种情况下，模型保存的概率估计给定个案取值 0 的概率，参数估计值应解释为与类别 0 的似然估计相关。
- 如果您指定第一个值（最低值）或 0 作为参考类别，则模型保存的概率估计给定个案取值 1 的概率。
- 如果您指定定制值并且变量已定义了标签，则可以通过从列表中选择值来设置参考类别。在指定模型的过程中并不确定某一特定变量的编码方式时，这种方法非常方便。

广义估计方程：预测变量

图片 7-5
广义估计方程：“预测变量”选项卡



在“预测”选项卡中可以指定用于生成模型效应的因子和协变量并指定可选偏移。

因子。因子是分类预测变量，可以是数值或字符串。

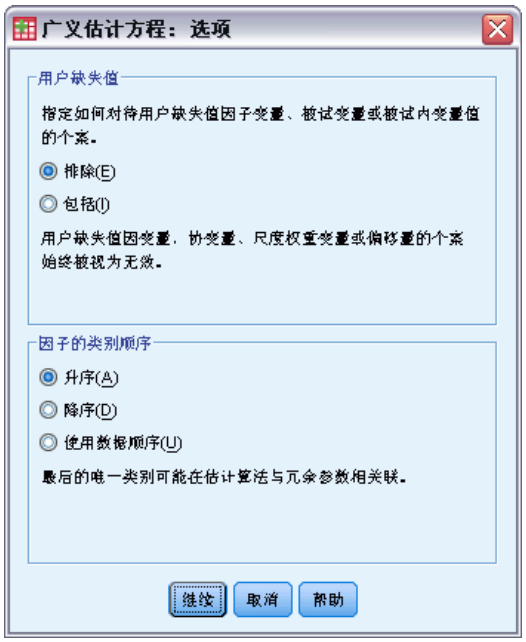
协变量。协变量为刻度预测变量，必须为数值。

注意：当响应为二元格式的二项式时，该过程通过子体来计算偏差和卡方拟合优度统计量，这些子体基于对所选因子和协变量观察值的交叉分类。您应在过程的多次运行中保持相同的预测变量集，以确保一致的子体数量。

偏移量。偏移项是“结构化”预测变量。模型不估计该预测变量的系数，但假定其值为 1；因此，偏移值只是简单地加到因变量的线性预测变量中。这在每个个案对于被观察事件都可能具有不同显现水平的泊松回归模型中尤其有用。例如，当对各个驾驶员事故率进行建模时，3 年驾驶经历出现 1 次事故和 25 年出现 1 次事故的驾驶员之间有着重大的差别！如果将驾驶员经历纳入偏移项，则事故数可以建模为泊松响应。

广义估计方程：选项

图片 7-6
“广义估计方程：选项”对话框



这些选项适用于“预测”选项卡上指定的所有因子。

用户缺失值。要在分析中包含个案，因子必须具有有效值。通过这些控制可以决定是否将用户缺失值在因子变量中视为有效值。

类别顺序。该选项与确定因子的最终水平有关，该水平可能与估计算法中的冗余参数相关。更改类别顺序可更改因子水平效应的值，因为这些参数估计值相对于“最终”水平进行计算。因子可以按从最低值到最高值的升序排序，按从最高值到最低值的降序排序，或者按“数据顺序”排序。这意味着在数据中遇到的第一个值定义第一个类别，遇到的最后一个唯一值定义最后一个类别。

广义估计方程：模型

图片 7-7
广义估计方程：“模型”选项卡

指定模型效应。缺省模型是仅截距模型，因此必须明确指定其他模型效应。还可以构建嵌套或非嵌套项。

非嵌套项

对于选定因子和协变量：

主效应。为每个选定的变量创建主效应项。

交互。为所有选定变量创建最高级交互项。

因子。创建选定变量的所有可能的交互和主效应。

所有二阶。创建选定变量的所有可能的二阶交互。

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。由于每位顾客只频繁光顾某一位置的商店，因此 Customer 效应可以说是**嵌套在** Store location 效应中。

此外，还可以包含交互效应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。

多项有序分布的模型没有单个截距项；而包含定义相邻类别之间转换点的阈值参数。模型中通常包含有阈值。

广义估计方程：估计

图片 7-8
广义估计方程：“估计”选项卡

参数估计。该组中的控件允许指定估计方法并提供参数估计值的初始值。

- **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，则算法将继续执行 Newton-Raphson 方法。
- **刻度参数方法。**可以选择尺度参数估计方法。

  最大似然法可联合估计尺度参数和模型效应；请注意，如果响应具有负二项式、泊松或二项式分布，则此选项无效。因为广义估计方程没有纳入似然这一概念，所以此规范仅适用于初始广义线性模型；该尺度参数估计值传递给广义估计方程，后者通过 Pearson 卡方与其自由度的商更新尺度参数。

  偏差和 Pearson 卡方选项从初始广义线性模型中统计量的值估计尺度参数；然后，此尺度参数估计值传递给广义估计方程，后者将其作为固定值处理。

或者，可为尺度参数指定固定值。在估计初始广义线性模型和广义估计方程时，该参数将作为固定值处理。

- **初始值。**该过程将自动计算参数的初始值。也可以指定参数估计值的初始值。

此选项卡上指定的迭代和收敛标准仅适用于初始广义线性模型。有关用于拟合广义估计方程的估计标准，请参见重复选项卡。

迭代。

- **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。
- **最大步骤对分。**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。
- **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。此选项可用于二元格式的多项式响应和二项式响应。

收敛性准则。

- **参数收敛。**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。
- **对数似然估计收敛。**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。
- **Hessian 收敛性。**对于“绝对值”指定，如果基于 Hessian 收敛的统计量小于指定的正值，则假定收敛。对于“相对”指定，如果统计量小于指定的正值和对数似然估计的绝对值的乘积，则假定收敛。

奇异性容许误差。奇异（非可逆）矩阵具有线性相关列，对估计算法可能产生严重问题。即使近似奇异的矩阵也可导致不良结果，因此该过程会将行列式小于容许误差的矩阵作为奇异矩阵对待。指定一个正值。

广义估计方程：初始值

该过程估计初始广义线性模型，从此模型得到的估计值用作广义估计方程的线性模型部分中参数估计值的初始值。处理相关矩阵无需初始值，因为矩阵元素基于参数估计值。此对话框上指定的初始值用作初始广义线性模型而不是广义估计方程的起始点，除非估计选项卡上的“最大迭代次数”设置为 0。

图片 7-9
“广义估计方程：初始值”对话框

如果指定了初始值，则必须为模型中的所有参数（包括冗余参数）提供初始值。在数据集中，变量从左到右的顺序必须为：RowType_、VarName_、P1、P2、…，其中 RowType_ 和 VarName_ 是字符串变量，P1、P2、… 是对应于已排序参数列表的数值变量。

- 初始值在变量 RowType_ 具有值 EST 的记录上提供；实际初始值在变量 P1、P2、… 下提供。该过程忽略 RowType_ 具有非 EST 值的所有记录以及 RowType_ 的第一次等于 EST 以外的所有记录。
- 截距（如果模型中包含）或阈值参数（如果响应具有多项式分布）必须是列出的第一个初始值。
- 尺度参数和负二项式参数（如果响应具有负二项式分布）必须是指定的最后一个初始值。
- 如果“拆分文件”有效，则变量必须以拆分文件变量开始（以创建“拆分文件”时指定的顺序），后跟 RowType_、VarName_、P1、P2、…，如上所述。拆分在指定数据集中的发生顺序必须与原始数据集中的顺序相同。

注意：变量名 P1、P2、… 不是必需的；过程接受参数的任何有效变量名，因为变量与参数的映射基于变量位置，而不是变量名。超过最后一个参数的所有变量都会被忽略。

初始值的文件结构与将模型导出为数据时使用的结构相同；因此，可以使用一次过程运行的最终值作为下一次运行的输入。

广义估计方程：统计量

图片 7-10
广义估计方程：“统计量”选项卡

模型作用。

- **分析类型。**指定要模型效应检验生成的分析类型。类型 I 分析通常适合于对模型中预测变量的排序具有先验理由的情况，而类型 III 应用范围更广。根据在“卡方统计量”组中的选择，计算 Wald 或一般化分数统计。
- **置信区间。**指定一个大于 50 且小于 100 的置信度。无论选择哪种卡方统计量类型，总是生成 Wald 区间，并且基于参数呈渐近正态分布的假设之上。
- **对数准似然函数。**它控制对数准似然函数的显示格式。整个函数包含一个相对于参数估计恒定的额外项；它对参数估计没有任何影响，并且在某些软件产品中不会显示。

打印。显示下面的输出。

- **个案处理摘要。**显示分析和“相关数据摘要”表中包含的个案和从中排除的个案的数量和百分比。
- **描述统计。**显示有关因变量、协变量和因子的描述统计和摘要信息。

- **模型信息。**显示数据文件名称、因变量或事件和试变量、偏移变量、尺度权重变量、概率分布和关联函数。
- **拟合度统计。** 显示用于模型选择的 AIC 准则的两项扩展：用于选择最佳相关结构的 QIC 准则和用于选择最佳预测变量子集的另一个 QIC 度量。
- **模型摘要统计。**显示模型拟合检验，包括用于模型拟合 Omnibus 检验的似然比统计量，以及每种效应的类型 I 或 III 对比的统计量。
- **参数估计。**显示参数估计值和相应的检验统计和置信区间。除了显示原始参数估计值之外，还可以选择显示取幂参数估计值。
- **参数估计的协方差矩阵。**显示估计参数协方差矩阵。
- **参数估计的相关性矩阵。**显示估计参数相关矩阵。
- **对比系数 (L) 矩阵。**为缺省效应和估计边际均值显示对比系数（如果“EM 均值”选项卡上要求）。
- **常规可估计函数。**显示生成对比系数 (L) 矩阵的矩阵。
- **迭代历史记录。**显示参数估计值和对数似然估计的迭代历史记录，并打印对梯度矢量和 Hessian 矩阵的最后一次评估。迭代历史记录表从第 0 次迭代（初始估计值）开始为每 n 次迭代显示参数估计值，其中，n 代表打印区间的值。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。
- **工作相关性矩阵。**显示表示主体内相关性的矩阵值。其结构取决于重复选项卡中的规格。

广义估计方程：EM 均值

图片 7-11
广义估计方程：“EM 均值”选项卡

此选项卡用于显示因子水平和因子交互的估计边际均值。您还可以要求显示整体估计均值。估计边际均值对有序多项模型不可用。

因子和交互。此列表包含“预测变量”选项卡上指定的因子和“模型”选项卡上指定的因子交互。此列表中不包含协变量。可以直接从此列表选择项，或者使用**依据 * **按钮将项组合到交互项中。

显示以下项的均值。计算所选因子和因子交互的估计均值。对比决定如何设定假设检验以比较估计均值。简单对比需要一个用于比较其他项的参考类别或因子水平。

- **配对方式。**为指定或默示的因子的所有水平组合计算成对比较。这是因子交互的唯一可用对比方法。
- **简单散点图.** 将每个水平的均值与指定水平的均值进行比较。当存在控制组时，此类对比很有用。

- **偏差。**因子的每个水平与总均值比较。偏移对比不是正交的。
- **差分.** 将每个水平（第一个除外）的均值与先前水平的均值进行比较。有时候将其称为逆 Helmert 对比。
- **Helmert.** 将因子的每个水平的均值（最后一个水平除外）与后续水平的均值进行比较。
- **重复.** 将每个水平的均值（最后一个水平除外）与后续水平的均值进行比较。
- **多项式.** 比较线性作用、二次作用、三次作用等等。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。

标度。对于响应，可根据因变量的原始刻度计算估算边际均值；对于线性预测变量，根据关联函数变换的因变量计算估计边际均值。

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。
- **Bonferroni.** 此方法针对检验多个对比这一事实调整观测的显著性水平。
- **连续 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 方法更严密的界限。
- **连续 Sidak.** 这是按顺序逐步降低的拒绝 Sidak 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。

广义估计方程：保存

图片 7-12
广义估计方程：“保存”选项卡

选中的项以指定的名称保存；可以选择用新变量覆盖现有的同名变量，或为新变量名添加后缀使其成为唯一名称以避免名称冲突。

- **响应均值的预测值。**以初始响应度规保存每个个案的模型预测值。当响应分布是二项式，并且因变量为二元时，过程会保存预测概率。当响应分布是多项式时，项目标签变成**累积预测概率**，并且过程保存响应每个类别的累积预测概率（最后一个除外），直至达到要保存的指定类别数量。
- **响应均值的置信区间下限**保存响应均值置信区间的下限。当响应分布是多项式时，项目标签变成**累积预测概率的置信区间下限**，并且过程保存响应每个类别的下限（最后一个除外），直至达到要保存的指定类别数量。
- **响应均值的置信区间上限。**保存响应均值置信区间的上限。当响应分布是多项式时，项目标签变成**累积预测概率的置信区间上限**，并且过程保存响应每个类别的上限（最后一个除外），直至达到要保存的指定类别数量。

- **预测类别。**对于具有二项式分布和二元因变量或多项式分布的模型，该过程保存每个个案的预测响应类别。此选项不支持其他响应分布。
- **线性预测器的预测值。**以线性预测变量的度规保存每个个案的模型预测值（通过指定的关联函数转换的响应）。当响应分布是多项式时，过程保存响应每个类别的预测值（最后一个除外），直至达到要保存的指定类别数量。
- **线性预测变量的预测值的估计标准误。**当响应分布是多项式时，过程保存响应每个类别的估计标准误（最后一个除外），直至达到要保存的指定类别数量。

如果响应分布是多项式，则以下各项不可用。

- **原始残差.** 观察值与模型预测值之间的差。
- **Pearson 残差。**个案对 Pearson 卡方统计量的贡献的平方根（带原始残差的符号）。

广义估计方程：导出

图片 7-13
广义估计方程：“导出”选项卡

将模型导出为数据。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、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。

- **拆分变量。**定义拆分的任何变量（如果使用）。
- **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）、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单独个案。
- **VarName_。** 对于行类型 COV 或 CORR，取值为 P1、P2、...，对应于所有估计模型参数（除尺度或负二项式参数外）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（必要时可包括尺度或负二项式参数）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、显著性水平和残差自由度设为系统缺失值。

  对于尺度参数，协方差、相关性、显著性水平和自由度均设为系统缺失值。如果尺度参数通过最大似然进行估计，则给出标准误；否则，它被设为系统缺失值。

  对于负二项式参数，协方差、相关性、显著性水平和自由度均设为系统缺失值。如果负二项式参数通过最大似然进行估计，则给出标准误；否则，它被设为系统缺失值。

  如果存在拆分，则参数列表必须在所有拆分之间进行累积。在给定拆分中，某些参数可能是无关的；但这不同于冗余。对于无关参数，所有协方差或相关性、参数估计值、标准误、显著性水平和自由度均设为系统缺失值。

可以使用此矩阵文件作为进一步进行模型估计的初始值；注意，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。尽管如此，您必须注意，在该矩阵文件中的所有参数对于读取此文件的过程具有相同的含义。

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。

GENLIN 命令的附加功能

使用命令语法语言还可以：

- 将参数估计值的初始值指定为数字列表（使用 CRITERIA 子命令）。
- 指定固定工作相关矩阵（使用 REPEATED 子命令）。
- 在计算估计边际均值时修正值不同于其均值的协变量（使用 EMMEANS 子命令）。
- 指定估计边际均值的定制多项式对比（使用 EMMEANS 子命令）。
- 指定因子的一个子集，以显示该子集的估计边际均值，并使用指定的对比类型对其进行比较（使用 EMMEANS 子命令的 TABLES 和 COMPARE 关键字）。

请参阅命令语法参考以获取完整的语法信息。

章

8

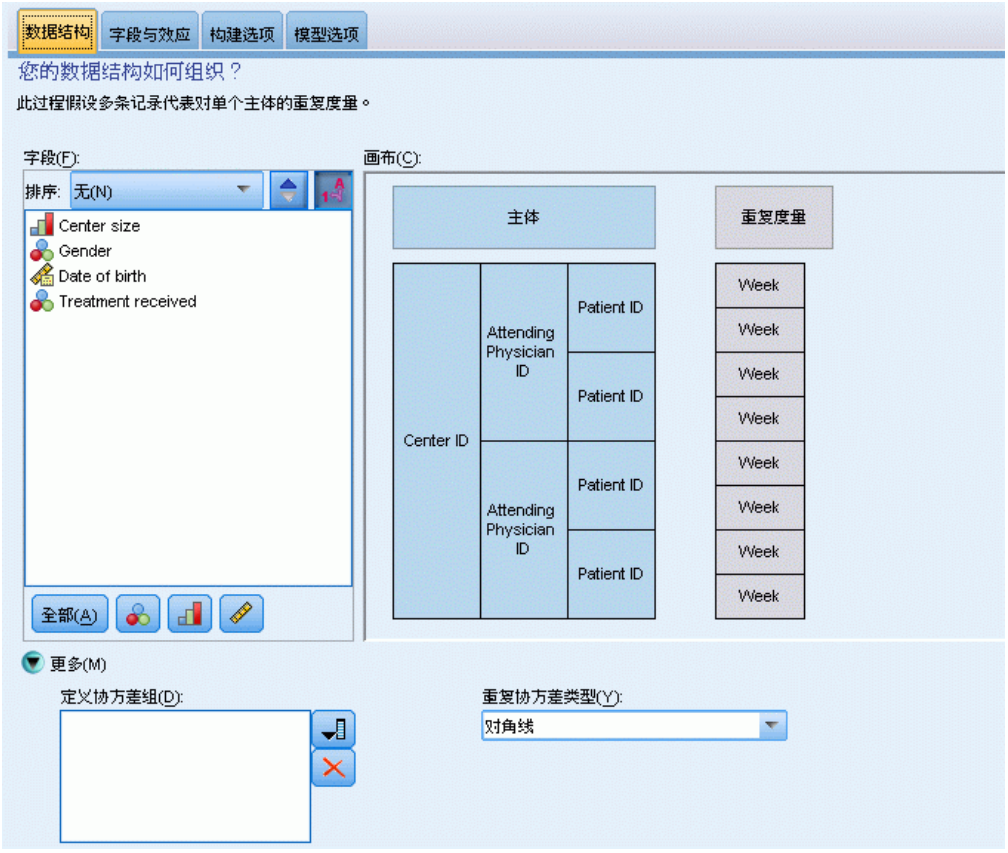
广义线性混合模型

广义线性混合模型扩展了线性模型，使得：

- 目标通过指定的关联函数与因子和协变量线性相关
- 目标可以有非正态分布
- 观测可能相关。

广义线性混合模型涵盖了从简单线性回归到复杂的非正态纵向数据多变量模型的各种模型。

图片 8-1
“数据结构”选项卡



“数据结构”选项卡用于在观察值相关时指定数据集中记录间的结构关系。如果数据集中的记录代表独立的观察值，则无需在此选项卡上指定任何设置。

主体。 指定分类字段的值组合应唯一定义数据集中的主体。例如，单个病人 ID 字段应足以定义一个医院内的主体，但如果病人标识号在医院间不唯一，则需要使用医院 ID 和病人 ID 的组合。在重复度量设置中，将为每个主体记录多个观察数据，因此每个主体可能在数据集中占用多个记录。

主体是可视为独立于其他主体的观察单元。例如，在医学研究中可以认为某患者的血压读数独立于其他患者的读数。如果存在对每个主体的重复度量，而且您想要对这些观察值之间的相关性建模，定义主体就非常重要。例如，您可能期望同一个患者在连续多次就医时得到的血压读数是相关的。

“数据结构”选项卡上被指定为主体的所有字段用于定义残差协方差结构的主体，并提供定义随机效应块上的随机效应协方差结构主体的可能字段列表。

重复度量。 这里指定的字段用于识别重复观测。例如，单个变量周可以标识医学研究中 10 周内的观察值，而月和天可共同用于标识一年内的每一天的观察值。

定义协方差组。这里指定的字段定义独立重复效应协方差参数组；每个类别各有一个参数组由分组字段的交叉分类定义。所有主体具有相同的协方差类型；相同协方差组内的主体具有相同的参数值。

重复协方差类型。这指定残差的协方差结构。可用的结构是：

- 一阶自回归 (AR1)
- 自回归移动平均 (1,1) (ARMA11)
- 复合对称
- 对角线
- 已标度的恒等
- Toeplitz
- 未结构化
- 方差成分

获取广义线性混合模型

此功能需要“高级回归”选项。

从菜单中选择：

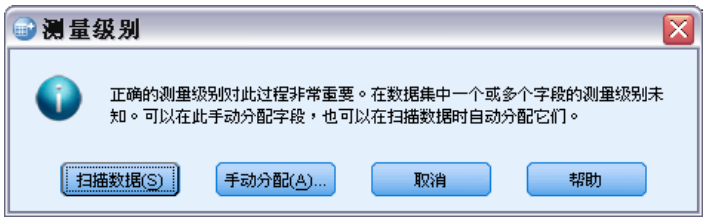
分析 > 混合模型 > 广义线性...

- 定义**数据结构**选项卡上数据集的主体结构。
- 在**字段和效应**选项卡上，必须有一个可具有任何测量级别的单个目标，或者是一个必须为连续的事件/试验指定。还可以指定其分布和关联函数、固定效应和任何随机效应块、偏移量或分析权重。
- 单击**构建选项**以指定可选的构建设置。
- 单击**模型选项**以保存得分到活动数据集并导出模型到外部文件。
- 单击**运行**以运行过程并创建模型对象。

具有未知测量级别的字段

当数据集中的一个或多个变量（字段）的测量级别未知时，将显示测量级别警告。由于测量级别会影响该过程的计算结果，因此所有变量必须都定义有测量级别。

图片 8-2
测量级别警报



- **扫描数据。** 读取活动数据集中的数据，并分配默认测量级别给任何具有当前未知测量级别的字段。如果数据集较大，该过程可能需要一些时间。
- **手动分配。** 打开列出了所有具有未知测量级别的字段的对话框。您可以使用该对话框将测量级别分配给这些字段。您也可以在数据编辑器的变量视图中分配测量级别。

由于测量级别对该过程很重要，因此您无法访问运行该过程的对话框，除非所有字段均定义了测量级别。

目标

图片 8-3
目标设置

这些设置通过关联函数定义目标、其分布以及其到预测变量的关系。

目标。 需要目标。它可具有任何测量级别，同时目标的测量级别可限制适当的分布和关联函数。

- **使用试验数作为分母。** 当目标响应是一系列试验中发生的事件数时，目标字段包含事件数，您可以额外选择一个包含试验数的字段。例如，在试验新型杀虫剂时，可以对蚂蚁样本施用不同浓度的杀虫剂，然后记录每个样本中杀灭的蚂蚁数量以及被施用杀虫剂的蚂蚁数量。在本例中，记录杀灭的蚂蚁数量的字段应指定为目标（事件）字段，记录每个样本中蚂蚁数量的字段应指定为试验字段。如果在每个样本中蚂蚁数量都相同，则可以将试验数指定为固定值。

  试验数应大于等于每个记录的事件数。事件应为非负整数，试验应为正整数。

- **自定义参考类别。** 对于分类目标，您可以选择参考类别。这会影响某些输出，如参数估计值，但不应更改模型拟合。例如，如果您的目标使用 0、1 和 2 的值，默认情况下，过程将最后（最高值）的类别或 2 作为参考类别。在这种情况下，参数估计应被解释为与类别 0 或 1 的似然相对于类别 2 的似然相关。如果您指定一个自定义类别，同时您的目标已定义标签，则可以通过从列表中选择一个值来设置参考类别。在指定模型的过程中并不确定某一特定字段的编码方式时，这种方法非常方便。

目标的分布以及与线性模型的关系（关联）。 给定预测变量的值，该模型将预期目标的值分布会遵循指定的形状，同时目标值通过指定关联函数与预测变量线性相关。可以使用现有的一些通用模型的快捷方式，或者如果您要拟合的特殊分布与关联函数组合不在此快捷列表上，还可以选择**自定义**设置。

- **线性模型。** 指定恒等关联的正态分布，在可使用线性回归或 ANOVA 模型预测目标时非常有用。
- **Gamma 回归。** 指定对数关联的 Gamma 分布，应在目标包含所有正值并偏斜到更大值时使用。
- **对数线性模型。** 指定对数关联的泊松分布，应在目标表示固定时间段内的发生计数时使用。
- **负二项式回归。** 指定对数关联的负二项式回归，应在目标和分母表示观察 k 次成功所需试验次数时使用。
- **多项 Logistic 回归。** 指定广义 Logit 关联的多项分布，应在目标是多类别响应时使用。
- **二元 logistic 回归。** 指定 logit 关联的二元分布，应在目标是通过 logistic 回归模型预测的二元响应时使用。
- **二元 probit。** 指定 probit 关联的二元分布，应在目标是基础正态分布的二元响应时使用。
- **间隔检查生存。** 指定互补双对数关联的二项式分布，在一些观测没有终止事件的生存分析中有用。

分布

此选项指定目标的分布。指定非正态分布和非恒等关联函数的功能是广义线性混合模型相对线性混合模型的重要改进。分布-关联函数可能存在多种组合，其中一些适合任何给定的数据集，因此可以根据先验理论的要求进行选择，或选择最合适的组合。

- **二项式。** 此分布仅适合表示二元响应或事件数量的目标。
- **Gamma。** 该分布适用于具有正尺度值并向更大的正值偏度的目标。如果数据值小于等于 0 或缺失，那么分析中不会使用相应的个案。
- **逆高斯。** 该分布适用于具有正尺度值并向更大的正值偏度的目标。如果数据值小于等于 0 或缺失，那么分析中不会使用相应的个案。
- **多项式。**该分布适用于表示多类别响应的目标。
- **负二项式。** 负二项回归使用带对数关联的二项分布，它用在目标代表具有较高方差的出现次数的情况下。
- **正态。** 该分布适合围绕某个中间值（均值）呈对称钟型分布的连续目标。
- **泊松。**该分布可视为被观察事件在固定时间段内发生的次数，适合具有非负整数值的变量。如果数据值是非整数、小于 0 或缺失，那么分析中不会使用相应的个案。

关联函数

关联函数是目标的转换形式，可用于模型估计。可用函数有：

- **恒等。**f(x)=x。不转换目标。该关联可用于除多项式以外的任何分布。
- **互补双对数。**f(x)=log(−log(1−x))。该函数只适用于二项式分布。

- **对数。**f(x)=log(x)。该关联可用于除多项式以外的任何分布。
- **对数补数。**f(x)=log(1−x)。该函数只适用于二项式分布。
- **Logit。**f(x)=log(x / (1−x))。该函数只适用于二项式或多项式分布，它也是多项式分布的唯一有效关联函数。
- **负双对数。**f(x)=−log(−log(x))。该函数只适用于二项式分布。
- **Probit。**f(x)=Φ^{-1}(x)，其中 Φ^{-1} 是逆标准正态累积分布函数。该函数只适用于二项式分布。
- **幂。**f(x)=x^{α}，如果 α ≠ 0. f(x)=log(x)，如果 α=0. α 为必需的数字指定，且必须为实数。该关联可用于除多项式以外的任何分布。

固定效应

图片 8-4
固定效应设置

固定效应因子通常被视为字段，其所需的值都表示在数据集中，同时可用于评分。默认情况下，在模型固定效应部分输入未在对话框中指定的具有预定义输入角色的字段。分类（名义、有序）字段用作模型中的因子，同时连续字段用作协变量。

通过在源列表中选择一个或多个字段并拖到效应列表来将效应输入至模型。所创建的效应类型取决于您放置选择的热区。

- **主。** 放置的字段在效应列表底部显示为单独的主效应。
- **二阶。** 放置字段的所有可能对在效应列表底部显示为二阶交互。
- **三阶。** 放置字段的所有可能三元在效应列表底部显示为三阶交互。
- ***。** 所有放置字段的组合在效应列表底部显示为单个交互。

效应构建器右侧的按钮可用于：

通过选择您要删除的项并单击删除按钮，可以从固定效应模型中删除项，

通过选择您要重新排序的项并单击向上或向下箭头，可以在固定效应模型中重新排序项，并

通过单击“添加自定义项”按钮，使用 添加自定义项 对话框向模型添加嵌套项。

包括截距。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。

添加自定义项

图片 8-5
添加“自定义项”对话框

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。由于每位顾客只频繁光顾某一位置的商店，因此 Customer 效应可以说是**嵌套在** Store location 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。

构建嵌套项

▶ 选择嵌套在另一个因子中的因子或协变量，然后单击箭头按钮。

▶ 单击（**内部**）。

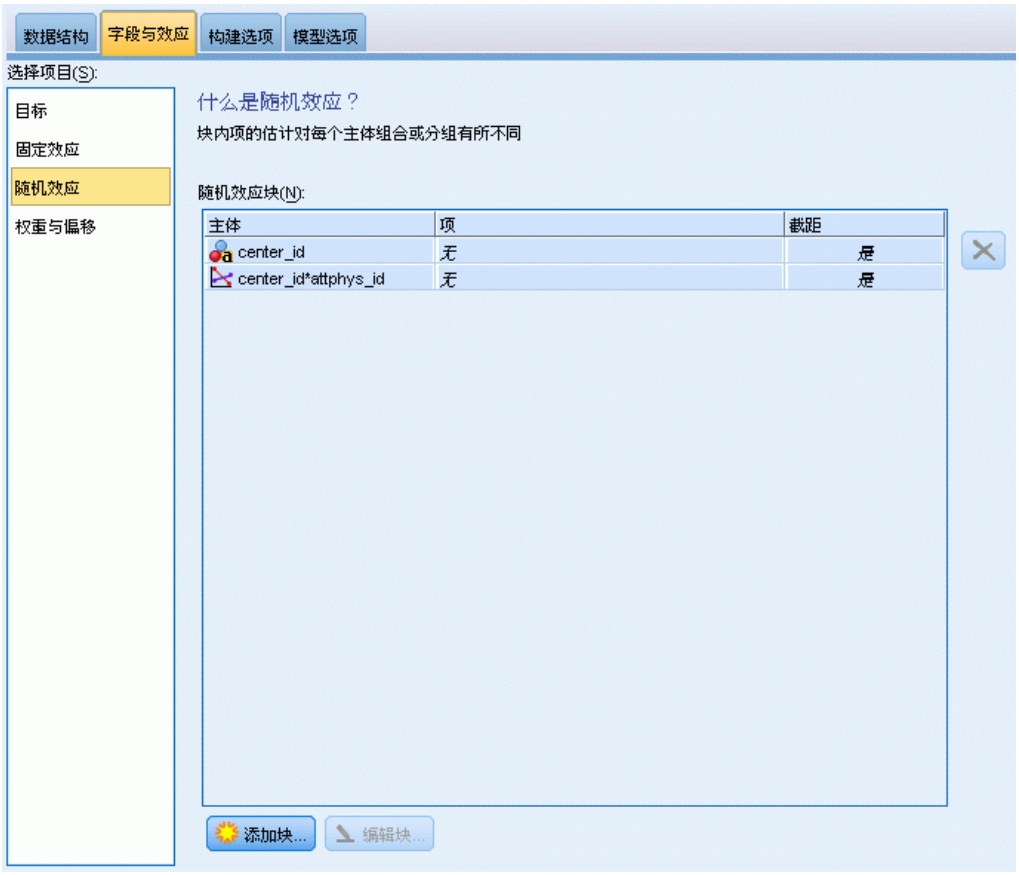
▶ 选择前一个因子或协变量嵌套在其中的因子，然后单击箭头按钮。

▶ 单击**添加项**。

（可选）可包含交互效应或者将多层嵌套添加到嵌套项中。

随机效应

图片 8-6
随机效应设置



随机效应因子是其在数据文件中的值可视为来自较大总体值的随机样本的字段。这对解释目标中的多余变异性十分有用。默认情况下，如果您在“数据结构”选项卡中选择了多个主体，将为超出最里面主体的每个主体创建一个随机效应块。例如，如果您选择“学校”、“班级”和“学生”作为“数据结构”选项卡上的主体，将自动创建以下随机效应块：

- 随机效应 1：主体是学校（没有效应，只有截距）
- 随机效应 2：主体是学校 * 班级（没有效应，只有截距）

您可以通过以下方式使用随机效应块：

- ▶ 要添加新块，单击**添加块...**这将打开 随机效应块 对话框。
- ▶ 要编辑现有块，选择您要编辑的块并单击**编辑块...**这将打开 随机效应块 对话框。
- ▶ 要删除一个或多个块，选择您要删除的块并单击删除按钮。

随机效应块

图片 8-7
“随机效应块”对话框

通过在源列表中选择一个或多个字段并拖到效应列表来将效应输入至模型。所创建的效应类型取决于您放置选择的热区。分类（名义、有序）字段用作模型中的因子，同时连续字段用作协变量。

- **主。** 放置的字段在效应列表底部显示为单独的主效应。
- **二阶。** 放置字段的所有可能对在效应列表底部显示为二阶交互。

- **三阶。** 放置字段的所有可能三元在效应列表底部显示为三阶交互。
- ***。** 所有放置字段的组合在效应列表底部显示为单个交互。

效应构建器右侧的按钮可用于：

通过选择您要删除的项并单击删除按钮，可以从固定效应模型中删除项，

通过选择您要重新排序的项并单击向上或向下箭头，可以在固定效应模型中重新排序项，并

通过单击“添加自定义项”按钮，使用 添加自定义项 对话框向模型添加嵌套项。

包括截距。默认情况下截距通常不包括在随机效应中。如果您可以假设数据穿过原点，则可以排除截距。

定义协方差组。这里指定的字段定义独立随机效应协方差参数组；分组字段的交叉分类定义的每个类别各有一个参数组。可以为每个随机效应块指定不同的分组字段集。所有主体具有相同的协方差类型；相同协方差组内的主体具有相同的参数值。

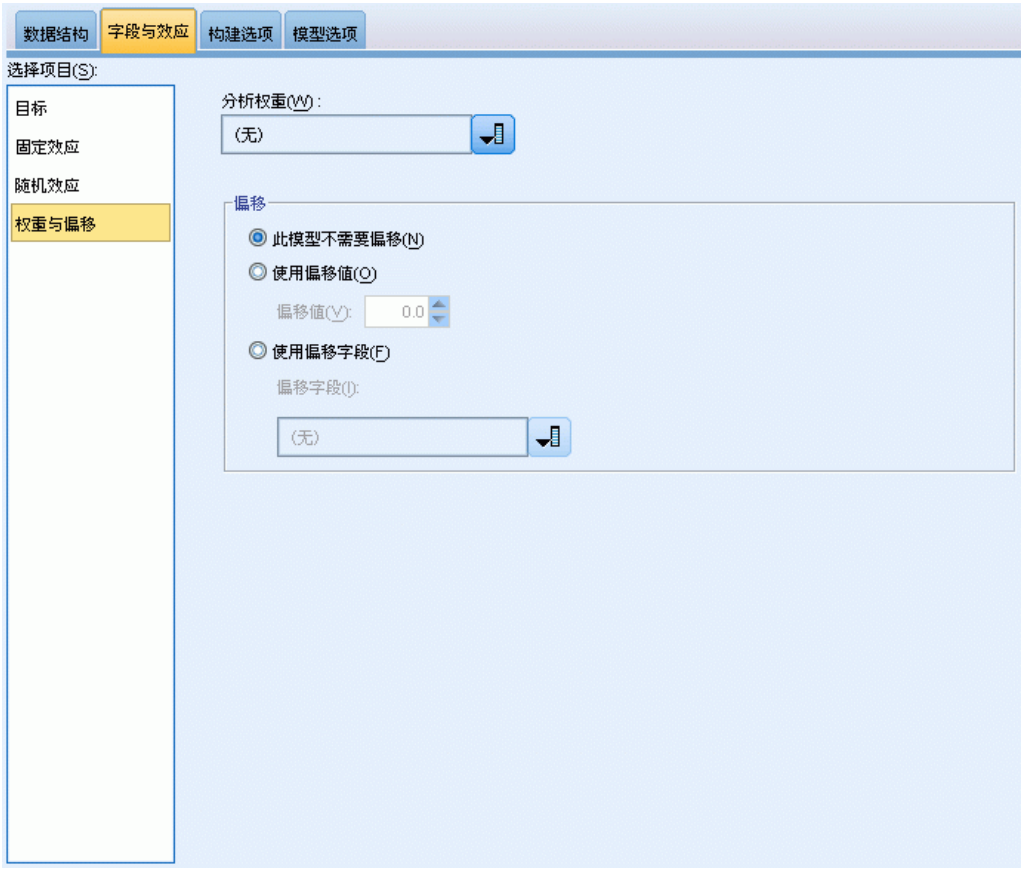
主体组合。您可以从“数据结构”选项卡的预设主体组合中指定随机效应主体。例如，如果学校、班级和学生定义为“数据结构”选项卡上的主体，同时以该顺序排序时，“主体”组合下拉列表将有**无**、**学校**、**学校 * 班级**和**学校 * 班级 * 学生**作为选项。

随机效应协方差类型。这指定残差的协方差结构。可用的结构是：

- 一阶自回归 (AR1)
- 自回归移动平均 (1,1) (ARMA11)
- 复合对称
- 对角线
- 已标度的恒等
- Toeplitz
- 未结构化
- 方差成分

权重和偏移量

图片 8-8
权重和偏移量设置



分析权重。尺度参数是与响应方差相关的估计模型参数。分析权重是“已知”值，可能因观察值的不同而异。如果指定了分析权重字段，则对每个观察值，都会用与响应方差相关的尺度参数除以分析权重值。分析中不使用分析权重值小于等于 0 或缺失的记录。

偏移量。偏移项是“结构化”预测变量。模型不估计该预测变量的系数，但假定其值为 1；因此，偏移值只是简单地加到目标的线性预测变量中。这在每个个案对于被观察事件都可能具有不同显现水平的泊松回归模型中尤其有用。例如，当对各个驾驶员事故率进行建模时，3 年驾驶经历出现 1 次事故和 25 年出现 1 次事故的驾驶员之间有着重大的差别！如果将驾驶员经历纳入偏移项，则事故数可以建模为泊松响应。

构建选项

图片 8-9
构建选项设置

这些选择指定一些用于构建模型的更高级条件。

排列顺序。 这些控件确定目标和因子（分类输入）的类别顺序，以确定“最后一个”类别。如果目标未分类或者在 目标 设置上指定了自定义参考类别，目标排序顺序设置将被忽略。

中止规则。 您可以指定算法将执行的最大迭代次数。指定一个非负整数。默认值为 100。

后估计设置。 这些设置确定如何计算一些用于查看的模型输出。

- **置信水平。** 这是用于计算模型系数的区间估计值的置信水平。指定一个大于 0 且小于 100 的值。默认值为 95。
- **自由度。** 这将指定计算显著性检验自由度的方法。如果您的样本大小足够大，或者数据非常平衡，或者模型使用比较简单的协方差类型（例如，已标度的恒等或对角线），选择**所有检验的自由度一样（残差方法）**。这是默认值。如果您的样本大小很小，或者数据不平衡，或者模型使用复杂的协方差类型（例如，未结构化），选择**根据检验不同而变化（Satterthwaite 近似值）**。
- **固定效应和系数检验** 这是计算参数估计协方差矩阵的方法。如果您担心破坏模型假设的话，选择稳健估计。

估算的均值

图片 8-10
估算的均值设置

此选项卡用于显示因子水平和因子交互的估计边际均值。估计边际均值对多项模型不可用。

项。 此处列出完全由分类字段组成的固定效应中的模型项。选中您希望模型生成估计边际均值的每个项。

- **对比类型。** 这将指定对比字段级别使用的对比类型。如果选择**无**，将不生成对比。**成对**为指定因子的所有水平组合生成成对比较。这是因子交互的唯一可用对比方法。**偏差**对比将因子的每个水平与总均值进行比较。**简单**对比将因子的每个水平（最后一个水平除外）与最后一个水平进行比较。“最后一个”水平由“构建选项”指定的因子排序顺序决定。注意，所有这些对比类别不正交。
- **对比字段。** 这将指定因子，其水平使用所选的对比类型进行比较。如果选择**无**作为对比类型，则无法（无需）选择任何对比字段。

连续字段。 列出的连续字段提取自使用连续字段的固定效应中的项。当计算估计边际均值时，协方差固定为指定值。选择均值或或指定自定义值。

选择估计均值。 这将指定是否基于目标原始尺度或关联函数转换计算估计边际均值。**原始目标尺度**计算目标的估计边际均值。注意，当使用事件/试验选项指定目标时，这将给出事件/试验比例而不是事件数量的估计边际均值。**关联函数转换**计算线性预测变量的估计边际均值。

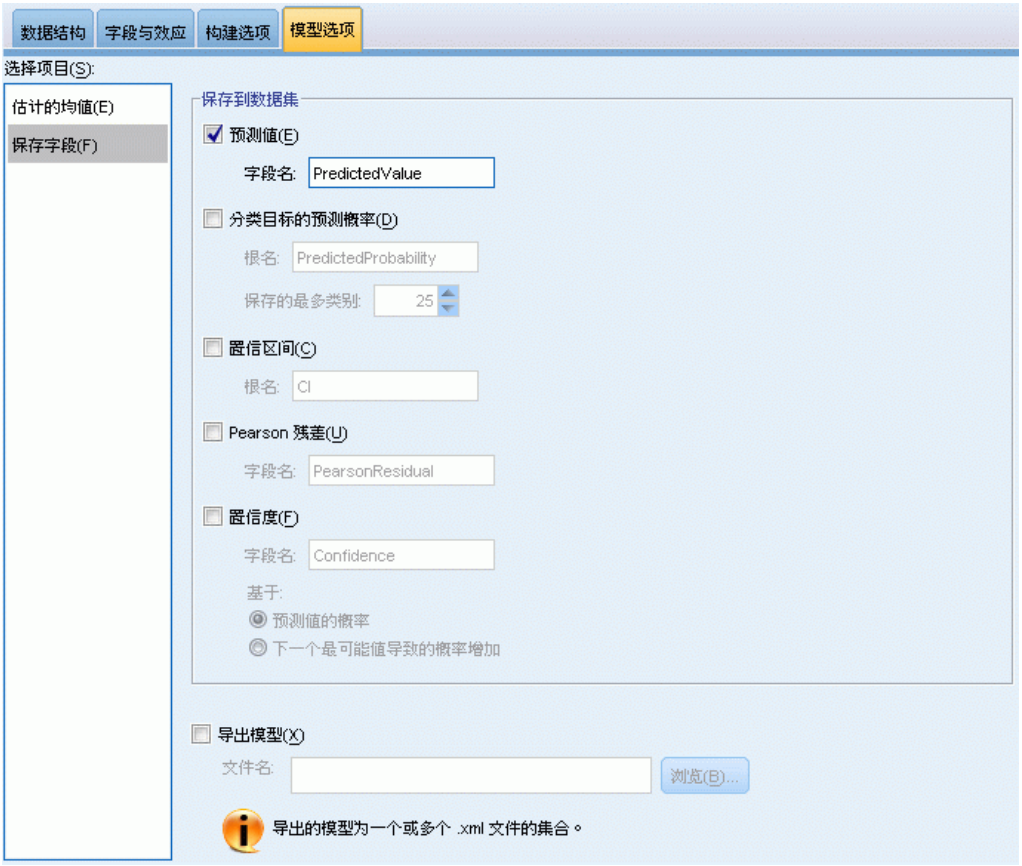
调整多重比较。在执行包含多重比较的假设检验时，总体显著性水平可从所包含的比较的显著性水平进行调节。这允许您选择调节方法。

- **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 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。
- **连续 Sidak.** 这是按顺序逐步降低的拒绝 Sidak 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。

最低显著性差异方法的保守程度不及顺序 Sidak 方法，后者的保守程度又不及顺序 Bonferroni 方法；这意味着，最低显著性差异会拒绝至少与顺序 Sidak 一样多的单个假设，而后者又会拒绝至少与顺序 Bonferroni 一样多的单个假设。

保存

图片 8-11
保存设置



选中项以指定名称保存；不允许与现有字段名称冲突。

预测值。保存目标的预测值。默认字段名称是 PredictedValue。

分类目标的预测概率。如果目标为分类，该关键字将保存前 n 个类别的预测概率，直到指定为**要保存的最大类别数**的值。默认根名称是 PredictedProbability。要保存预测类别的预测概率，可以保存置信水平（参阅下文）。

置信区间。 保存预测值或预测概率的置信区间的上限和下限。对于除多项式之外的所有分布，这将创建两个变量，同时默认根名称是 CI，后缀为 _Lower 和 _Upper 。

对于多项式分布，为最后一个类别（）之外的每个自变量类别创建一个字段。这将为前 n 个类别保存累积预测概率的下限和上限，直到但不包括最后一个列表，以及直到指定为**要保存的最大类别数**的值。默认根名称是 CI，同时默认字段名称是 CI_Lower_1、CI_Upper_1、CI_Lower_2、CI_Upper_2，依此类推，与目标类别的顺序对应。

Pearson 残差。为每个记录保存 Pearson 残差，该残差可用于模型拟合的后估计诊断。默认字段名称是 PearsonResidual。

置信。 以预测值保存分类目标的置信。计算置信可基于预测值的概率（最高的预测概率）或最高预测概率和其次高预测概率之间的差异。默认字段名称是 Confidence。

导出模型。这将写入模型到外部 .zip 文件。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。 指定有效的唯一文件名。如果文件规范引用了现有文件，那么该文件将被覆盖。

模型视图

此过程在查看器中创建“模型”对象。激活（双击）该对象，可获得模型的交互式视图。

默认情况下，显示“模型摘要”视图。要查看另一个模型视图，从视图缩略图中选择。

模型摘要

该视图是模型及其拟合的快照概览摘要。

表。该表识别目标设置上指定的目标、概率分布和关联函数。 如果目标由事件和试验定义，单元格将被拆分以显示事件字段和试验字段，或试验的固定次数。此外，将显示经有限样本校正的 Akaike 信息准则 (AICC) 和 Bayesian 信息准则 (BIC)。

- **Akaike 校正.** 选择和比较基于 -2（约束）对数似然估计的混合模型的度量。值越小，模型越好。AICC“修正”适用于小样本的 AIC。随着样本尺寸增加，AICC 收敛为 AIC。
- **Bayesian.** 选择和比较基于 -2对数似然估计的模型的度量。值越小，模型越好。BIC 也“惩罚”参数过多的模型，且比 AIC 更严格。

图表。如果目标为分类，将有一张图表显示最终模型的精确性，即正确分类的百分比。

数据结构

该视图提供您指定的数据结构的摘要，并帮助您检查是否正确指定主体和重复度量。将为每个主体字段和重复度量字段以及目标显示所观察到的第一个主体信息。此外，将显示每个主体字段和重复度量字段的级别数。

按已观测进行预测

对于连续目标，包括指定为事件/试验的目标，这将显示一个分级散点图，其中预测值位于垂直轴上，而观测值位于水平轴上。理想情况下，该点应在 45 度线上；您可以从该视图上判断出任何被模型预测为较差的纪录。

分类

对于分类目标，将以热图显示已观测和已预测值的交叉分类，以及整体正确百分比。

表样式。有几个不同的显示样式，可以从**样式**下来列表中访问这些样式。

- **行百分比。**这将在单元格中显示行百分比（单元计数以行总计百分比表示）。这是默认值。
- **单元格计数。**这将显示单元格中的单元格计数。热图的阴影还是基于行百分比。
- **热图。**这将在单元格中不显示任何值，只显示阴影。
- **压缩。**这将在单元格中不显示任何行或列标题，或值。当目标有许多类别时，它将非常有用。

缺失。如果目标上的任何记录缺失值，则会显示在所有有效行下的 （**缺失**）行中。具有缺失值的记录不会对整体正确百分比作出贡献。

多个目标。如果有多个分类目标，那么每个目标将显示在单个表中，同时有一个**目标**下拉列表控制要显示的目标。

大型表。如果所显示的目 有 100 多个类别，将不显示任何表。

固定效应

此视图显示模型中每个固定效应的大小。

样式。有多种不同的显示样式，可以从**样式**下拉列表中访问这些样式。

- **图表。**这是按照效应在“固定效应”设置上指定的顺序从顶部到底部对它们进行排序的图表。在图表中，连接线条根据效应的显著性进行加权，粗线条表示较显著的效应（p 值较小）。这是默认值。
- **表**此为总体模型与单独模型效应的 ANOVA 表。单个效应按照它们在“固定效应”设置上指定的顺序从顶部到底部进行排序。

显著性。这里有一个“显著性”滑块，以控制在视图中显示哪些效应。显著性值大于滑块值的效应将被隐藏。这不会改变模型，只是帮助您重点关注最重要的效应。默认情况下此值为 1.00，因此不会根据显著性来过滤效应。

固定系数

此视图显示模型中每个固定系数的值。注意，由于因子（分类预测变量）在模型内部经过指示符编码，因此包含因子的**效应**通常具有多个关联**系数**；每种类别一个关联系数，但对应于冗余系数的类别除外。

样式。有多种不同的显示样式，可以从**样式**下拉列表中访问这些样式。

- **图表。**这是首先显示截距，然后按照效应在“固定效应”设置上指定的顺序从顶部到底部对它们进行排序的图表。在包含因子的效应中，系数按照数据值的升序进行排列。在图表中，连接线条根据系数的显著性进行加权并具有不同颜色，粗线条表示较显著的系数（p 值较小）。这是默认样式。
- **表**这将显示单独模型系数的值、显著性检验，以及置信区间。截距后，效应按照它们在“固定效应”设置上指定的顺序从顶部到底部进行排序。在包含因子的效应中，系数按照数据值的升序进行排列。

多项式。如果多项分布有效，那么多项式列表将控制要显示的目标类别。值在列表中的排序顺序由“构建选项”设置上的指定决定。

指数分布。这将显示某些模型类型的指数系数估计和置信区间，包括二元 logistic 回归（二元分布和 logit 关联）、名义 Logistic 回归（多项分布和 logit 关联）、负二项式回归（负二项式分布和对数关联）和 Log-linear 模型（泊松分布和对数关联）。

显著性。这里有一个“显著性”滑块，以控制在视图中显示哪些系数。显著性值大于滑块值的系数将被隐藏。这不会改变模型，只是帮助您重点关注最重要的系数。默认情况下此值为 1.00，因此不会根据显著性来过滤系数。

随机效应协方差

该视图显示随机效应协方差矩阵（G）。

样式。有多种不同的显示样式，可以从**样式**下拉列表中访问这些样式。

- **协方差值。**这是按照效应在“固定效应”设置上指定的顺序从顶部到底部对它们进行排序的协方差矩阵热图。相关图中的颜色对应于单元值，如键中所示。这是默认值。
- **相关图。** 这是协方差矩阵的热图。
- **压缩。**这是不带行和列标题的协方差矩阵热图。

块。如果有多个随机效应块，那么将有一个“块”下拉列表来选择要显示的块。

组。 如果随机效应块有组指定，那么将有一个“组”下拉列表来选择要显示的组级别。

多项式。如果多项分布有效，那么多项式列表将控制要显示的目标类别。值在列表中的排序顺序由“构建选项”设置上的指定决定。

协方差参数

该视图显示残差和随机效应的协方差参数估计和相关统计量。这些高级但基本的结果可提供协方差结构是否合适的信息。

摘要表。这是对残差（R）和随机效应（G）协方差矩阵中的参数数量、固定效应（X）和随机效应（Z）设计矩阵中的秩（列数）以及定义数据结构的主体字段所定义的主体数量的快速参考。

协方差参数表。对于选定的效应，将为每个协方差参数显示估计、标准误和置信区间。所显示的参数数量取决于效应的协方差结构，同时对于随机效应块，则取决于块中的效应数量。如果您看到非对角线参数不显著，您可以使用更加简单的协方差结构。

效应。如果有多个随机效应块，那么将有一个“块”下拉列表来选择要显示的残差或随机效应块。残差效应总是可用。

组。 如果残差或随机效应块有组指定，那么将有一个“组”下拉列表来选择要显示的组级别。

多项式。如果多项分布有效，那么多项式列表将控制要显示的目标类别。值在列表中的排序顺序由“构建选项”设置上的指定决定。

估算的均值：显著效应

这些是显示 10 个“最显著”固定所有因子效应的图表，以三阶交互开始，然后是二阶交互，最后是主效应。该图表会为水平轴上主效应（或交互中首个列出的效应）的每个值在垂直轴上显示目标模型估计值；为交互中的第二个列出效应的每个值生成单独的线；为三阶交互中的第三个列出效应的每个值生成单独的图表；所有其他预测变量保持恒定。它提供了有关每个预测变量系数在目标上的效应的直观表示，非常有用。注意，如果没有显著的预测变量，则不会生成估计均值。

置信。这将使用指定为“构建选项”一部分的置信水平显示边际均值的置信上限和下限。

估算的均值：自定义效应

这些是用于用户请求的固定所有因子效应的表和图表。

样式。有多种不同的显示样式，可以从**样式**下拉列表中访问这些样式。

- **图表。**该样式显示水平轴上主效应（或交互中首个列出的效应）的每个值在垂直轴上的目标模型估计值线图；为交互中的第二个列出效应的每个值生成单独的线；为三阶交互中的第三个列出效应的每个值生成单独的图表；所有其他预测变量保持恒定。

  如果请求了对比，将显示另一个图表以比较对比字段的水平；对于交互，将为对比字段之外的每个效应水平组合显示图表。对于**成对**比较，它是距离网络图，即比较表的图形表示，其中网络中节点间的距离对应于样本间的差别。黄线对应于统计上的显著差异；黑线对应于不显著的差异。悬停在网络中的直线上将显示具有该直线连接的节点间的调整差异显著性的工具提示。

  对于**偏差**对比会显示条形图，垂直轴显示目标模型估计值，水平轴上显示对比字段值；对于交互，将为对比字段之外的每个效应水平组合显示图表。这些条形图将以黑色水平线表示对比字段的每个水平和整体均值之间的差异。

  对于**简单**对比会显示条形图，垂直轴显示目标模型估计值，在水平轴上显示对比字段值；对于交互，将为对比字段之外的每个效应水平组合显示图表。这些条形图将以黑色水平线表示对比字段每个水平（最后一个水平除外）和最后一个水平之间的差异。

- **表。**该样式显示目标模型估计值、其标准误和效应中每个字段水平组合的置信区间；所有其他预测变量保持恒定。

  如果请求了对比，另一个表将显示每个对比的估计、标准误、显著性检验和置信区间；对于交互，除对比字段之外，效应的每个效应水平组合都有一组单独的行。此外，将显示带整体检验结果的表；对于交互除了对比字段之外，效应的每个效应水平组合都有单独的整体检验。

置信。这将切换使用指定为“构建选项”一部分的置信水平的边际均值的置信上限和下限的显示。

布局。这将切换成对对比图表的布局。圆形布局揭示的对比比网络布局更少，但会避免重叠线。

章
9

模型选择对数线性分析

“模型选择对数线性分析”过程分析多阶交叉制表（列联表）。它使用成比例拟合的迭代算法将分层对数线性模型拟合到多维交叉制表。此过程可帮助您找出关联的分类变量。要构建模型，可以使用强制输入和向后去除方法。对于饱和模型，可以请求参数估计值和偏关联检验。饱和模型会为所有单元加上 0.5。

示例。在研究两种洗涤剂中的一种的用户偏好时，研究人员统计了每组的人数、水的软硬度（软、中等或硬）的各种类别、其中一个品牌的上一次使用以及洗涤温度（冷或热）。他们发现了温度与水的软硬度以及品牌偏好的关系。

统计量。频率、残差、参数估计值、标准误、置信区间和偏关联检验。对于定制模型，则为残差图和正态概率图。

数据。因子变量是分类的。要分析的所有变量都必须是数值。开始进行模型选择分析前，可以将分类字符串变量重新编码为数值变量。

避免指定具有多个水平的多个变量。这样的指定可能导致多个单元具有少量的观察值，卡方值可能没用。

相关过程。“模型选择”过程可帮助标识模型中需要的项。然后您就可以使用“一般对数线性分析”或“Logit 对数线性分析”继续评估模型。可以使用“自动重新编码”重新编码字符串变量。如果数值变量具有空类别，则使用“重新编码”创建连续的整数值。

获取模型选择对数线性分析

从菜单中选择：

分析 > 对数线性 > 模型选择...

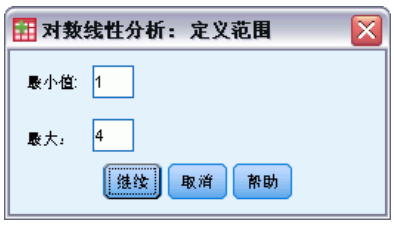
图片 9-1
“模型选择对数线性分析”对话框

- 选择两个或更多数值型分类因子。
- 在“因子”列表中选择一个或多个因子变量，然后单击**定义范围**。
- 定义每个因子变量的值范围。
- 在“建立模型”组中选择一个选项。

  或者，您可以选择一个单元权重变量指定结构中的无效单元。

对数线性分析：定义范围

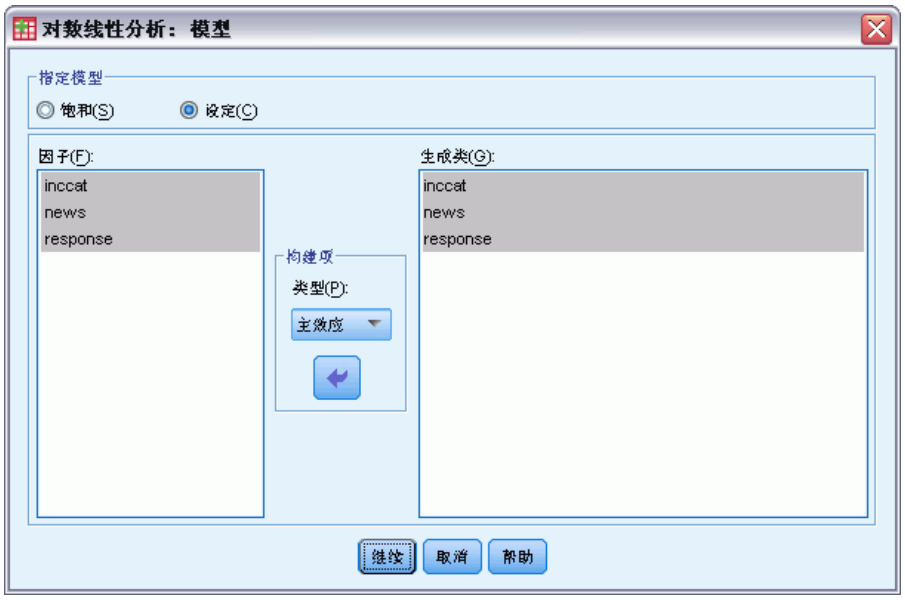
图片 9-2
“对数线性分析：定义范围”对话框



必须指明每个因子变量的类别的范围。“最小值”和“最大值”的值与因子变量的最低和最高类别相对应。这两个值都必须是整数，并且最小值必须小于最大值。将排除值在边界以外的个案。例如，如果您指定最小值为 1，最大值为 3，则只使用 1、2 和 3 这三个值。对每个因子变量重复这一过程。

对数线性分析：模型

图片 9-3
“对数线性分析：模型”对话框



指定模型。饱和模型包含所有因子的主效应以及所有因子与因子的交互作用。选择**定制**可为不饱和模型指定生成类。

生成类。生成类是因子出现的最高阶项的列表。一个分层模型包含定义生成类和所有低阶相关性的项。假设您在“因子”列表中选择了变量 A、B 和 C，然后在“构建项”下拉列表中选择了**交互**。生成的模型将包含指定的三阶交互 A*B*C，二阶交互 A*B、A*C 和 B*C，以及 A、B 和 C 的主效应。不要在生成类中指定低阶相关性。

构建项

对于选定因子和协变量：

交互。创建所有选定变量的最高级交互项。这是缺省值。

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二阶交互。

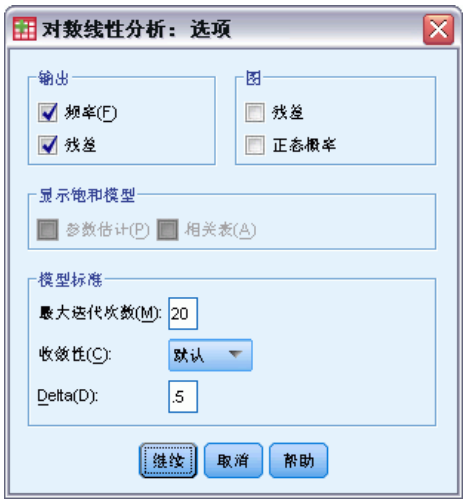
所有三阶。创建选定变量的所有可能的三阶交互。

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。

模型选择对数线性分析：选项

图片 9-4
“对数线性分析：选项”对话框



显示。您可以选择频率和/或残差。在饱和模型中，观察的和期望的频率相同，残差为 0。

图。对于定制模型，可以选择两种类型的图中的一种或两种：残差和正态概率。这些可帮助确定模型与数据的拟合度。

显示饱和模型。对于饱和模型，可以选择参数估计值。参数估计值可帮助确定可从模型中删除哪一项。此外还有一个可用的关联表，其中列出了偏关联检验。对于具有多个因子的表，选择这个选项需要进行大量的计算。

模型标准。使用成比例拟合的迭代算法获取参数估计值。通过指定最大迭代次数、收敛或 Delta（为饱和模型的所有单元频率添加的值）可覆盖一个或多个估计标准。

HILOGLINEAR 命令的附加功能

使用命令语法语言还可以：

- 以矩阵形式指定单元权重（使用 CWEIGHT 子命令）。
- 使用单个命令生成对多个模型的分析（使用 DESIGN 子命令）。

请参见命令语法参考以获取完整的语法信息。

章
10

一般对数线性分析

“一般对数线性分析”过程分析落入交叉制表或列联表中每个交叉分类类别的观察值频率计数。表中的每个交叉分类构成一个单元，每个分类变量称为一个因子。因变量为交叉制表单元中的个案数（频率），解释变量为因子和协变量。此过程使用 Newton-Raphson 方法估计分层和非分层对数线性模型的最大似然参数。可以分析泊松或多项分布。

您最多可以选择 10 个因子来定义表的单元。单元结构变量允许定义不完整表的结构中的无效单元，在模型中包含偏移项，拟合对数比率模型或实现边际表的调整方法。对比变量允许计算广义对数几率比 (GLOR)。

模型信息和拟合优度统计量自动显示。还可以显示各种统计量和图，或在活动数据集中保存残差和预测值。

示例。使用佛罗里达汽车事故报告中的数据可以确定系上安全带与致命还是非致命伤势之间的关系。几率比指示关系的显著证据。

统计量。观察的和期望的频率；原始残差、调整残差和偏差残差；设计矩阵；参数估计值；几率比；对数几率比；GLOR；Wald 统计；以及置信区间。图：调整残差、偏差残差和正态概率。

数据。因子是分类变量，单元协变量是连续变量。当模型中有协变量时，会将单元中个案的协变量均值应用于该单元。对比变量是连续的。它们用于计算广义对数几率比。对比变量的值是期望单元计数的对数线性组合的系数。

单元结构变量指定权重。例如，如果一些单元是结构中的无效单元，则单元结构变量值为 0 或 1。请勿使用单元结构变量对分类汇总数据加权。而应从“数据”菜单选择**加权个案**。

假设。一般对数线性分析提供两种分布：泊松和多项分布。

在泊松分布假设下：

- 总样本大小在研究前不固定，或分析不取决于总样本大小。
- 在单元中出现观察的事件在统计上独立于其他单元的单元计数。

在多项分布假设下：

- 总样本大小是固定的，或分析取决于总样本大小。
- 单元计数在统计上不是独立的。

相关过程。使用“交叉表”过程检查交叉制表。如果应将一个或多个分类变量视为响应变量而将其他变量视为解释变量，则使用“Logit 对数线性”过程。

获取常规对数线性分析

▶ 从菜单中选择：

分析 > 对数线性 > 一般...

图片 10-1
“一般对数线性分析”对话框

▶ 在“常规对数线性分析”对话框中，选择最多 10 个因子变量。

根据需要，您可以：

- 选择单元协变量。
- 选择单元结构变量以定义结构中的无效单元或包含偏移项。
- 选择对比变量。

一般对数线性分析模型

图片 10-2
“一般对数线性分析：模型”对话框

指定模型。饱和模型包含涉及因子变量的所有主效应和交互效应。它不包含协变量项。选择**定制**可以仅指定其中一部分的交互或指定因子协变量交互。

因子与协变量。列出因子与协变量。

模型中的项。模型取决于数据的性质。选择**定制**之后，您可以选择分析中感兴趣的主效应和交互效应。必须指定要包含在模型中的所有项。

构建项

对于选定因子和协变量：

交互。创建所有选定变量的最高级交互项。这是缺省值。

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二阶交互。

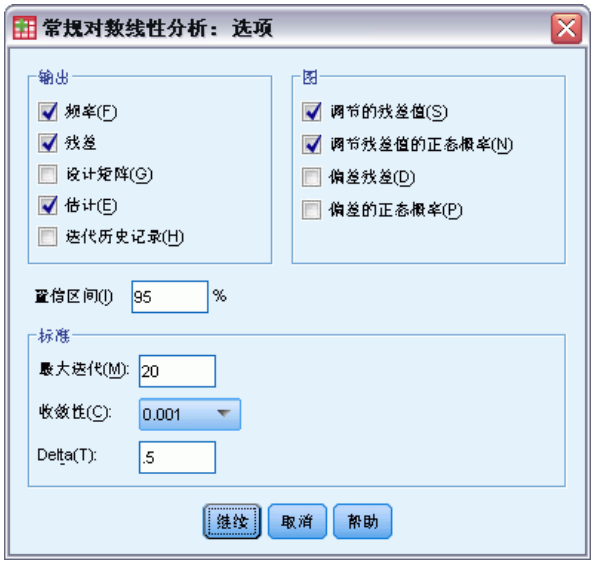
所有三阶。创建选定变量的所有可能的三阶交互。

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。

一般对数线性分析：选项

图片 10-3
“一般对数线性分析：选项”对话框



“一般对数线性分析”过程显示模型信息和拟合优度统计量。此外，还可以选择以下选项中的一个或多个：

显示。多个统计量可供显示：观察到的单元格频率和期望单元格频率；原始残差、调整残差和偏差残差；模型设计矩阵；以及模型的参数估计值。

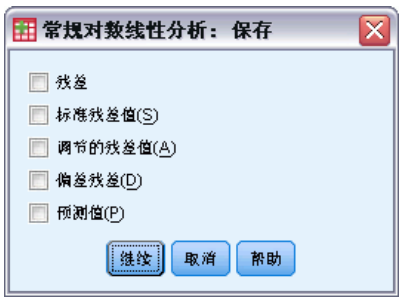
图。图只能用于定制模型，其中包含两个散点图矩阵（针对观察的单元计数和期望的单元计数的调整残差或偏差残差）。还可以显示调整残差或偏差残差的正态概率和反趋势正态图。

置信区间。可以调整参数估计值的置信区间。

标准。Newton-Raphson 方法用于获取最大似然参数估计值。可以为最大迭代次数、收敛标准和 delta（添加到所有初始近似值单元的常数）输入新值。Delta 保留在饱和模型的单元中。

一般对数线性分析：保存

图片 10-4
“一般对数线性分析：保存”对话框



选择要保存为活动数据集中的新变量的值。新变量名称中的后缀 n 会递增，以使每个保存变量都具有唯一的名称。

保存值指分类汇总数据（列联表中的单元），即使数据记录在数据编辑器的各个观察值中。如果为未分类汇总的数据保存残差或预测值，则将列联表中单元的保存值输入到该单元中每个个案的数据编辑器中。要使保存值有意义，应分类汇总数据以获得单元计数。

可以保存四种类型的残差：原始残差、标准化残差、调整残差和偏差残差。还可以保存预测值。

- **残差**. 也称为简单残差或原始残差，它是观察单元计数及其期望计数之间的差分。
- **标准残差值**. 残差除以其标准误的估计。标准化残差也称为 Pearson 残差。
- **调节的残差值**. 标准残差值除以其估计的标准误。由于当选定模型正确时，调整残差为渐近标准正态分布，因此它们在检查正态性方面优于标准化残差。
- **偏差残差**. 个体对似然比卡方统计量的贡献的带符号平方根（G 的平方），其中的符号是残差的符号（观察到的计数减去期望计数）。偏差残差具有渐近标准正态分布。

GENLOG 命令的附加功能

使用命令语法语言还可以：

- 计算观察到的单元频率和期望单元频率的线性组合，并打印该组合的残差、标准化残差和调整残差（使用 GERESID 子命令）。
- 更改冗余检查的缺省阈值（使用 CRITERIA 子命令）。
- 显示标准化残差（使用 PRINT 子命令）。

请参见命令语法参考以获取完整的语法信息。

章
11

Logit 对数线性分析

“Logit 对数线性分析”过程分析因变量（或响应变量）与自变量（或解释变量）之间的关系。因变量始终为分类变量，而自变量可以是分类变量（因子）。其他自变量、单元协变量可以是连续变量，但它们不在逐个案的基础上应用。单元的加权协变量均值应用于该单元。因变量几率的对数表示为参数的线性组合。自动采用多项分布；这些模型有时称为多项 Logit 模型。此过程使用 Newton-Raphson 算法估计 Logit 对数线性模型的参数。

您一共可以选择 1 到 10 个因变量与因子变量。单元结构变量允许定义不完整表的结构中的无效单元，在模型中包含偏移项，拟合对数比率模型或实现边际表的调整方法。对比变量允许计算广义对数几率比 (GLOR)。对比变量的值是期望单元计数的对数线性组合的系数。

模型信息和拟合优度统计量自动显示。还可以显示各种统计量和图，或在活动数据集中保存残差和预测值。

示例。佛罗里达的一项研究包含了 219 条鳄鱼。鳄鱼的食物类型如何随它们的体形大小和它们居住的四个湖而改变？研究发现，体形较小的鳄鱼选择爬虫类代替鱼类为食的几率是体形较大的鳄鱼的 0.70 倍；并且主要选择爬虫类代替鱼类为食的几率在第 3 个湖中最高。

统计量。观察的和期望的频率；原始残差、调整残差和偏差残差；设计矩阵；参数估计值；广义对数几率比；Wald 统计；和置信区间。图：调整残差、偏差残差和正态概率图。

数据。因变量是分类变量。因子是分类变量。单元协变量可以是连续的，但当模型中有协变量时，会将单元中个案的协变量均值应用于该单元。对比变量是连续的。它们用于计算广义对数几率比 (GLOR)。对比变量的值是期望单元计数的对数线性组合的系数。

单元结构变量指定权重。例如，如果一些单元是结构中的无效单元，则单元结构变量值为 0 或 1。请勿使用单元结构变量对分类汇总数据加权。而应使用“数据”菜单中的“加权个案”。

假设。假设解释变量的每个类别组合中的计数具有多项分布。在多项分布假设下：

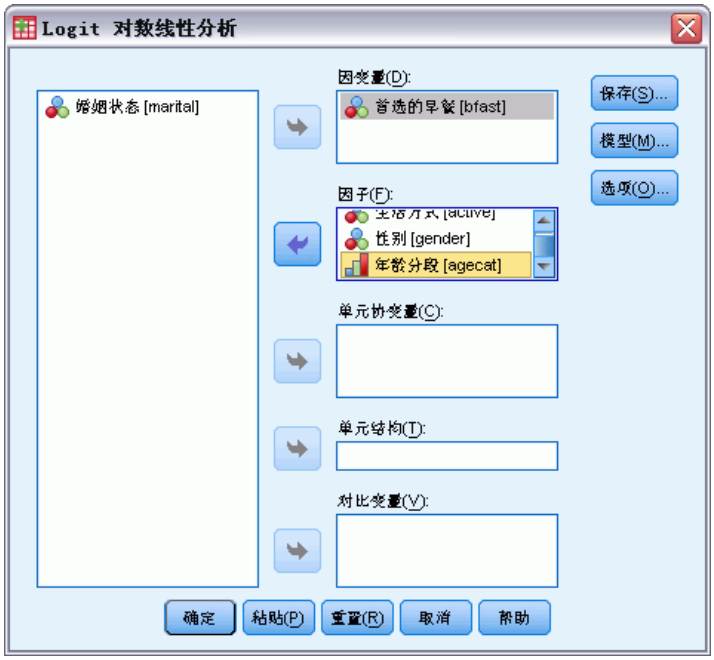
- 总样本大小是固定的，或分析取决于总样本大小。
- 单元计数在统计上不是独立的。

相关过程。使用“交叉表”过程显示列联表。当希望分析观察到的计数和一组解释变量之间的关系时，请使用“一般对数线性分析”过程。

获取 Logit 对数线性分析

▶ 从菜单中选择：
分析 > 对数线性 > Logit...

图片 11-1
“Logit 对数线性分析”对话框



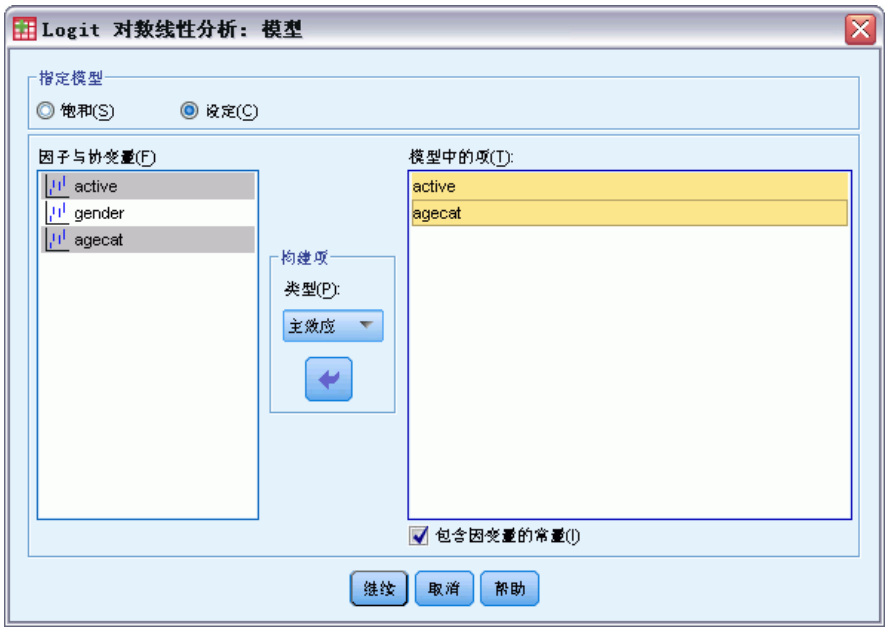
- 在“Logit 对数线性分析”对话框中，选择一个或多个因变量。
- 选择一个或多个因子变量。

  因变量和因子变量的总数必须小于或等于 10。

  根据需要，您可以：
  - 选择单元协变量。
  - 选择单元结构变量以定义结构中的无效单元或包含偏移项。
  - 选择一个或多个对比变量。

Logit 对数线性分析：模型

图片 11-2
“Logit 对数线性分析：模型”对话框



指定模型。饱和模型包含涉及因子变量的所有主效应和交互效应。它不包含协变量项。选择**定制**可以仅指定其中一部分的交互或指定因子协变量交互。

因子与协变量。列出因子与协变量。

模型中的项。模型取决于数据的性质。选择**定制**之后，您可以选择分析中感兴趣的主效应和交互效应。必须指定要包含在模型中的所有项。

通过取因变项的所有可能组合并将每个组合与模型列表中的每项匹配，将所有项添加到设计中。如果选择了**将常数包含为因变量**，则还会将一个单元项 (1) 添加到模型列表中。

例如，假设变量 D1 和 D2 是因变量。“Logit 对数线性分析”过程创建一个因变项列表 (D1, D2, D1*D2)。如果“模型中的项”列表包含 M1 和 M2 并且包含一个常数，则模型列表包含 1、M1 和 M2。生成的设计包括每个模型项与每个因变项的组合：

D1、D2、D1*D2

M1*D1、M1*D2、M1*D1*D2

M2*D1、M2*D2、M2*D1*D2

将常数包含为因变量。在定制模型中将常数包含为因变量。

构建项

对于选定因子和协变量：

交互。创建所有选定变量的最高级交互项。这是缺省值。

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二阶交互。

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。

Logit 对数线性分析：选项

图片 11-3
“Logit 对数线性分析：选项”对话框

“Logit 对数线性分析”过程显示模型信息和拟合优度统计量。此外，还可以选择以下选项中的一个或多个：

显示。 多个统计量可供显示：观察到的单元格频率和期望的单元格频率；原始残差、调整残差和偏差残差；模型设计矩阵；以及模型的参数估计值。

图。用于定制模型的图包含两个散点图矩阵（针对观察到的单元计数和期望单元计数的调整残差或偏差残差）。还可以显示调整残差或偏差残差的正态概率和反趋势正态图。

置信区间。可以调整参数估计值的置信区间。

标准。Newton-Raphson 方法用于获取最大似然参数估计值。可以为最大迭代次数、收敛标准和 delta（添加到所有初始近似值单元的常数）输入新值。Delta 保留在饱和模型的单元中。

Logit 对数线性分析：保存

图片 11-4
“Logit 对数线性分析：保存”对话框

选择要保存为活动数据集中的新变量的值。新变量名称中的后缀 n 会递增，以使每个保存变量都具有唯一的名称。

保存值指分类汇总数据（列联表中的单元），即使数据记录在数据编辑器的各个观察值中。如果为未分类汇总的数据保存残差或预测值，则将列联表中单元的保存值输入到该单元中每个个案的数据编辑器中。要使保存值有意义，应分类汇总数据以获得单元计数。

可以保存四种类型的残差：原始残差、标准化残差、调整残差和偏差残差。还可以保存预测值。

- **残差**. 也称为简单残差或原始残差，它是观察单元计数及其期望计数之间的差分。
- **标准残差值**. 残差除以其标准误的估计。标准化残差也称为 Pearson 残差。
- **调节的残差值**. 标准残差值除以其估计的标准误。由于当选定模型正确时，调整残差为渐近标准正态分布，因此它们在检查正态性方面优于标准化残差。
- **偏差残差**. 个体对似然比卡方统计量的贡献的带符号平方根（G 的平方），其中的符号是残差的符号（观察到的计数减去期望计数）。偏差残差具有渐近标准正态分布。

GENLOG 命令的附加功能

使用命令语法语言还可以：

- 计算观察到的单元频率和期望单元频率的线性组合，并打印该组合的残差、标准化残差和调整残差（使用 GERESID 子命令）。
- 更改冗余检查的缺省阈值（使用 CRITERIA 子命令）。
- 显示标准化残差（使用 PRINT 子命令）。

请参见命令语法参考以获取完整的语法信息。

章
12

寿命表

在多数情况下，您都会希望考察两个事件之间的时间分布，比如雇用时长（员工从雇用到离开公司的时间）。但是，这类数据通常包含没有记录其第二次事件的个案（例如，在调查结束后仍然为公司工作的员工）。这种情况的发生有以下几个原因：对于某些个案，事件在研究结束前没有发生；而对于另一些个案，我们在研究结束前的某段时间未能跟踪其状态；还有一些个案可能因一些与研究无关的原因（例如员工生病或请假）无法继续。这些个案总称为**已审查的个案**，它们使得此类研究不适合 t 检验或线性回归等传统方法。

用于此类数据的统计方法称为跟进**寿命表**。寿命表的基本概念是将观察区间划分为较小的时间区间。对于每个区间，使用所有观察至少该时长的人员计算该区间内发生期间终结的概率。然后使用从每个区间估计的概率估计在不同时间点发生该事件的整体概率。

示例。新尼古丁贴片疗法是否比传统贴片疗法更有助于戒烟？您可以对两组吸烟者进行调查，一组接受传统疗法，另一组接受实验性疗法。从数据构造寿命表将允许您比较两组的整体戒烟率，以确定实验性疗法是否是传统疗法的改进。还可以用图来表示生存或风险函数并对其进行直观比较，以获得更详细的信息。

统计量。每组在每个时间区间的期初记入数、期末离开数、历险数、期间终结数、终结比例、生存比例、累积生存比例（和标准误）、概率密度（和标准误）以及风险率（和标准误）；每组的中位数生存时间；用于比较两组间生存分布的 Wilcoxon (Gehan) 检验。图：生存、对数生存、密度、风险率和 1 减生存的函数图。

数据。时间变量应是定量的。状态变量应是以整数编码的二分变量或分类变量，事件编码为单值或一段连续值范围。因子变量应是以整数编码的分类变量。

假设。所关心事件的概率应只取决于初始事件之后的时间（假设绝对时间下的概率不变）。即，从不同时间开始研究的个案（比如，从不同时间开始接受治疗的患者）应有相似的行为。已审查的个案和未审查的个案之间也不应存在系统性差别。例如，如果许多已审查的个案都是情况更为严重的患者，则得到的结果可能会存在偏差。

相关过程。“寿命表”过程对此类分析（通常称为“生存分析”）使用保险精算方法。“Kaplan-Meier 生存分析”过程使用略有不同的方法计算寿命表，此方法不依赖于将观察期划分为较小的时间区间。如果您的观察值数量较少，建议使用此方法，这样每个生存时间区间内将只有较少数量的观察值。如果您怀疑变量与要控制的生存时间或变量（协变量）相关，则应使用“Cox 回归”过程。如果同一个个案中协变量在不同的时间点可以具有不同的值，则应使用带有“依时协变量”的“Cox 回归”。

创建寿命表

▶ 从菜单中选择：
分析 > 生存 > 寿命表...

图片 12-1
“寿命表”对话框

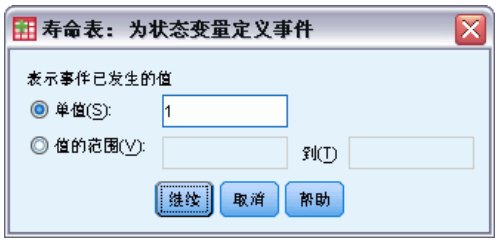
- 选择一个**数值**生存变量。
- 指定要检查的时间区间。
- 选择一个状态变量定义已发生期间终结的个案。
- 单击**定义事件**指定用于指示事件发生的状态变量的值。

  或者，可以选择一阶因子变量。会为因子变量的每个类别生成生存变量的保险精算表。

  还可以选择二阶按因子变量。会为一阶和二阶因子变量的每个组合生成生存变量的保险精算表。

寿命表：为状态变量定义事件

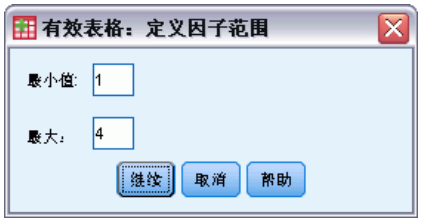
图片 12-2
“寿命表：为状态变量定义事件”对话框



为状态变量选择的一个或多个值的出现指示这些个案已发生期间终结。所有其他个案视为已审查。输入标识感兴趣事件的单值或值范围。

寿命表：定义范围

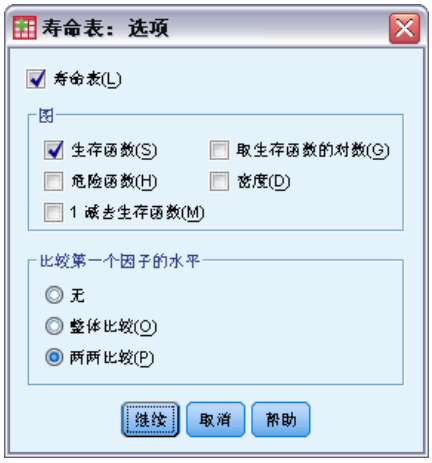
图片 12-3
“寿命表：定义范围”对话框



因子变量的值在所指定范围内的个案将包括在分析中，并会为该范围内的每个唯一值生成单独的表（和图，如果要求）。

寿命表：选项

图片 12-4
“寿命表：选项”对话框



您可以控制寿命表分析的各个方面。

寿命表。若要在输出中不显示寿命表，请取消选择**寿命表**。

图。允许您请求生存函数图。如果已经定义了因子变量，则会为因子变量定义的每个子组生成图。可用图包括生存、对数生存、风险、密度和 1 减生存。

- **生存**. 在线性刻度上显示累积生存函数。
- **取生存函数的对数**. 在对数刻度上显示累积生存函数。
- **危险函数**. 在线性刻度上显示累积风险函数。
- **密度**. 显示密度函数。
- **1 减去生存函数**. 以线性尺度绘制 1 减生存函数。

比较第一个因子的水平。如果有一阶控制变量，则可以在此组中选择一个选项执行 Wilcoxon (Gehan) 检验，该检验比较子组生存。检验对一阶因子执行。如果已经定义了二阶因子，则会对二阶变量的每个水平执行检验。

SURVIVAL 命令的附加功能

使用命令语法语言还可以：

- 选择多个因变量。
- 指定间距不等的区间。
- 选择多个状态变量。
- 指定不包含所有因子和所有控制变量的比较。
- 计算近似而不是精确比较。

请参见命令语法参考以获取完整的语法信息。

章
13

Kaplan-Meier 生存分析

在多数情况下，您都会希望考察两个事件之间的时间分布，比如雇用时长（员工从雇用到离开公司的时间）。但是，这种数据通常包含一些已审查的个案。已审查的个案是没有记录其第二次事件的个案（例如，在调查结束后仍然为公司工作的员工）。Kaplan-Meier 过程是已审查的个案出现时估计时间事件模型的一种方法。Kaplan-Meier 模型的依据是估计事件发生的每个时间点的条件概率，并取这些概率的乘积限估计每个时间点的生存率。

示例。新的 AIDS 疗法在延长寿命方面是否具有治疗优势？您可以对两组 AIDS 患者进行研究，一组接受传统疗法，另一组接受实验性疗法。从数据构造 Kaplan-Meier 模型将允许您比较两组的整体生存率，以确定实验性疗法是否是传统疗法的改进。还可以用图来表示生存或风险函数并对其进行直观比较，以获得更详细的信息。

统计量。生存表，包括时间、状态、累积生存和标准误、累积事件和剩余数；以及均值和中位数生存时间，带有标准误和 95% 置信区间。图：生存、风险、对数生存和 1 减生存。

数据。时间变量应为连续变量，状态变量可以是分类变量或连续变量，因子和层次变量应为分类变量。

假设。所关心事件的概率应只取决于初始事件之后的时间（假设绝对时间下的概率不变）。即，从不同时间开始研究的个案（比如，从不同时间开始接受治疗的患者）应有相似的行为。已审查的个案和未审查的个案之间也不应存在系统性差别。例如，如果许多已审查的个案都是情况更为严重的患者，则得到的结果可能会存在偏差。

相关过程。Kaplan-Meier 过程使用的计算寿命表的方法估计每个事件发生时的生存或风险函数。“寿命表”过程使用保险精算方法进行生存分析，该方法依赖于将观察期划分为较小的时间区间，可能对处理大样本有用。如果您怀疑变量与要控制的生存时间或变量（协变量）相关，则应使用“Cox 回归”过程。如果同一个个案中协变量在不同的时间点可以具有不同的值，则应使用带有“依时协变量”的“Cox 回归”。

获取 Kaplan-Meier 生存分析

▶ 从菜单中选择：
分析 > 生存 > Kaplan-Meier...

图片 13-1
“Kaplan-Meier”对话框

- 选择时间变量。
- 选择一个状态变量标识已发生期间终结的个案。该变量可以是数字或**短字符串**。然后单击**定义事件**。

  或者，可以选择因子变量检查组差异。还可以选择对变量的每个水平（层次）生成单独分析的层次变量。

Kaplan-Meier：为状态变量定义事件

图片 13-2
“Kaplan-Meier: 为状态变量定义事件”对话框

输入表示已出现期间终结的值。可以输入单个值、值范围或值列表。只有在状态变量为数值时，“值范围”选项才可用。

Kaplan-Meier： 比较因子水平

图片 13-3
“Kaplan-Meier: 比较因子水平”对话框

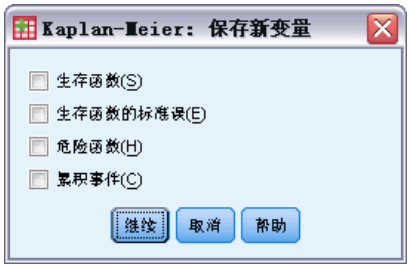
您可以请求统计量以检验因子不同水平的生存分布的等同性。可用统计量包括对数秩、Breslow 和 Tarone-Ware。选择一个选项指定要进行的比较：跨层整体检验、分层检验、跨层成对检验或分层成对检验。

- **对数秩.** 比较生存分布的等同性的检验。在此检验中，所有时间点均赋予相同的权重。
- **Breslow.** 比较生存分布的等同性的检验。在每个时间点用带风险的个案数对时间点加权。
- **Tarone-Ware.** 比较生存分布的等同性的检验。在每个时间点用历险的个案数的平方根对时间点加权。
- **在层上比较所有因子水平.** 在单次检验中比较所有因子水平，以检验生存曲线的相等性。
- **在层上成对比较因子水平.** 比较每一个相异的因子水平对。不提供成对趋势检验。
- **对于每层.** 对每层的所有因子水平的相等性执行一次单独的检验。如果您没有分层变量，则不执行检验。
- **为每层成对比较因子水平.** 比较每一层的每一个相异的因子水平对。不提供成对趋势检验。如果您没有分层变量，则不执行检验。

因子级别的线性趋势。允许您检验跨因子级别的线性趋势。此选项仅可用于因子水平的整体（而不是成对）比较。

Kaplan-Meier： 保存新变量

图片 13-4
“Kaplan-Meier: 保存新变量”对话框

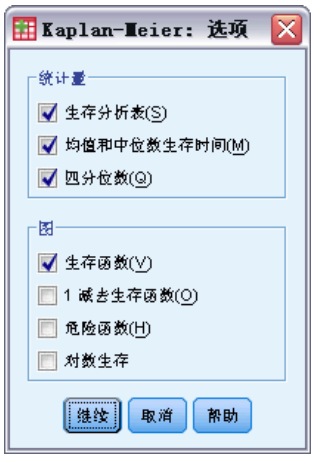


您可以将 Kaplan-Meier 表的信息保存为新变量，新变量可在以后的分析中用于检验假设或检查假设。您可以将生存函数、生存函数的标准误、危险函数和累积事件保存为新变量。

- **生存**. 累积生存概率估计。默认变量名为前缀 sur_ 加上顺序号。例如，如果已存在 sur_1，Kaplan-Meier 就分配变量名 sur_2。
- **生存函数的标准误**. 累积生存估计的标准误。默认变量名为前缀 se_ 加上顺序号。例如，如果已存在 se_1，Kaplan-Meier 就分配变量名 se_2。
- **危险函数**. 累积风险函数估计。默认变量名为前缀 haz_ 加上顺序号。例如，如果已存在 haz_1，Kaplan-Meier 就分配变量名 haz_2。
- **累积事件**. 当个案按其生存时间和状态代码进行排序时的事件累积频率。默认变量名为前缀 cum_ 加上顺序号。例如，如果已存在 cum_1，Kaplan-Meier 就分配变量名 cum_2。

Kaplan-Meier：选项

图片 13-5
“Kaplan-Meier: 选项”对话框



您可以从 Kaplan-Meier 分析请求多种输出类型。

统计量。您可以选择为计算的生存函数显示统计量，包括生存分析表、均值和中位数生存时间以及四分位数。如果包含因子变量，则会为每组生成单独的统计量。

图。通过图可以直观地检查生存函数、1 减去生存函数、危险函数和取生存函数的对数。如果包含因子变量，则会为每组绘制函数图。

- **生存**. 在线性刻度上显示累积生存函数。
- **1 减去生存函数**. 以线性尺度绘制 1 减生存函数。
- **危险函数**. 在线性刻度上显示累积风险函数。
- **取生存函数的对数**. 在对数刻度上显示累积生存函数。

KM 命令的附加功能

使用命令语法语言还可以：

- 获得将要追踪调查的个案损失作为独立于已审查个案的单独类别的频率表。
- 为线性趋势检验指定不等间距。
- 获取生存时间变量的百分位数而不是四分位数。

请参见命令语法参考以获取完整的语法信息。

章
14

Cox 回归分析

Cox 回归为时间事件数据建立预测模块。该模块生成生存函数用于为预测变量的给定值预测被观察事件在给定时间内 t 发生的概率。从观察主体中估计预测的生存函数形状与回归系数；该方法稍后可应用于具有预测变量测量的新个案。注意，已检查主体中的信息，即未在观察时间内经历被观察事件的信息，为模块估计做出巨大贡献。

示例。男性和女性因吸烟引发肺癌的风险是否不同？通过构造一个 Cox 回归模型，输入吸烟情况（每天吸烟根数）和性别作为协变量，您可以检验关于性别和吸烟情况对肺癌发作的影响的假设。

统计量。对于每个模型：－2LL，似然比统计量和整体卡方。对于模型中的变量：参数估计值、标准误和 Wald 统计量。对于不在模型中的变量：得分统计量和残差卡方。

数据。时间变量应是定量变量，但状态变量可以是分类或连续变量。自变量（协变量）可以是连续或分类变量；如果是分类变量，它们应经过哑元编码或指示符编码（该过程中有一个自动对分类变量进行编码的选项）。层次变量应是分类变量，编码为整数或短字符串。

假设。观察值应是独立的，风险比应是时间恒定值；即，各个个案风险的比例不应随时间变化。后一个假设称为**成比例的风险假设**。

相关过程。如果成比例的风险假设不成立（请参见上文），可能需要使用带依时协变量的 Cox 过程。如果没有协变量或者只有一个分类协变量，可以使用寿命表或 Kaplan-Meier 过程检查样本的生存或风险函数。如果样本中没有已审查的数据（即，每个个案都出现期间终结），可以使用线性回归过程对预测变量和时间事件之间的关系进行建模。

获得 Cox 回归分析

▶ 从菜单中选择：
分析 > 生存函数 > Cox 回归...

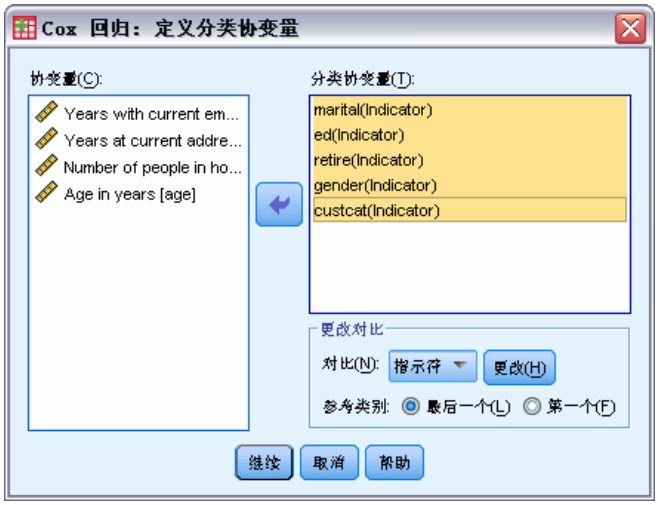
图片 14-1
“Cox 回归”对话框

- 选择时间变量。不分析时间值为负值的个案。
- 选择一个状态变量，然后单击**定义事件**。
- 选择一个或多个协变量。要包含交互项，请选择交互中所涉及的所有变量，然后单击 >a*b>。

  或者，可以通过定义分层变量为不同组计算各自的模型。

Cox 回归：定义分类变量

图片 14-2
“Cox 回归：定义分类协变量”对话框



您可以详细指定 Cox 回归过程处理分类变量的方式。

协变量。列出在主对话框中指定的所有协变量，无论是直接指定的协变量还是作为交互的一部分在任何层中指定的协变量。如果其中部分协变量是字符串变量或分类变量，则能将它们用作分类协变量。

分类协变量。列出标识为分类变量的变量。每个变量都在括号中包含一个表示法，指示要使用的对比编码。字符串变量（由变量名称后的符号 < 指示）已存在于“分类协变量”列表中。从“协变量”列表中选择其他任意分类协变量并将它们移到“分类协变量”列表中。

更改对比。可用于更改对比方法。可用的对比方法有：

- **指示符。**这些对比指示类别成员资格是否存在。参考类别在对比矩阵中表示为一排“0”。
- **简单。**除参考类别外，预测变量的每个类别都与参考类别相比较。
- **差分。**除第一个类别外，预测变量的每个类别都与前面的类别的平均效应相比较。也称为逆 Helmert 对比。
- **Helmert。**除最后一个类别外，预测变量的每个类别都与后面的类别的平均效应相比较。
- **重复。**除第一个类别外，预测变量的每个类别都与它前面的那个类别进行比较。
- **多项式。**正交多项式对比。假设类别均匀分布。多项式对比仅适用于数值变量。
- **偏差。**除参考类别外，预测变量的每个类别都与总体效应相比较。

如果选择**偏差**、**简单**或**指示符**，则可以选择**第一个**或**最后一个**作为参考类别。注意，直到单击**更改**后，该方法才实际发生更改。

字符串协变量必须是分类协变量。要从“分类协变量”列表中移去某字符串变量，必须从主对话框中的“协变量”列表中移去所有包含该变量的项。

Cox 回归：图

图片 14-3
“Cox 回归：图”对话框

图有助于评估估计的模型和解释结果。您可以对生存函数、危险函数、负对数累积生存函数的对数和 1 减去生存函数绘图。

- **生存函数**. 在线性刻度上显示累积生存函数。
- **危险函数(H)**. 在线性刻度上显示累积风险函数。
- **对数减对数。** 向估计应用了 ln（-ln）转换之后的累积生存估计。
- **1 减去生存函数(M)**. 以线性尺度绘制 1 减生存函数。

因为这些函数依赖于协变量的值，所以必须对协变量使用常数值来绘制函数与时间的关系图。默认情况是使用每个协变量的平均值作为常数值，但可以使用“更改值”控制组输入您自己的值用于绘图。

可以将分类协变量移入“对应的各条线”文本框，从而为该协变量的每个值单独绘制一条线。此选项仅对分类协变量可用，分类协变量在“协变量值的绘制位置”列表中由名称后的 (Cat) 指示。

Cox 回归：保存新变量

图片 14-4
“Cox 回归：保存新变量”对话框

可以将分析的各种结果保存为新变量。可以在以后的分析中使用这些变量来检验假设或检查假设。

保存模型变量。允许您将回归的生存函数及其标准误、对数负对数估计、风险函数、偏残差和 DfBeta，以及线性预测变量 X*Beta 保存为新变量。

- **函数(F)**. 给定时间的累积剩余函数值。该值等于生存到那个时间段概率。
- **负对数累积生存函数的对数(L)**. 在将 ln(-ln) 变换应用于估计之后的累积生存估计。
- **危险函数(H)**. 保存累积风险函数估计（又称为 Cox-Snell 残差）。
- **偏残差(P)**. 您可以对照生存时间来绘制偏残差，以检验成比例的风险假定。为最终模型中的每个协变量保存一个变量。仅对包含至少一个协变量的模型提供偏残差。
- **DfBeta(B)**. 在剔除了某个个案的情况下系数的估计更改。为最终模型中的每个协变量保存一个变量。仅对包含至少一个协变量的模型提供 DfBetas。
- **X*Beta**. 线性预测变量得分。每个个案的以均值为中心的协变量值及其对应的参数估计值的乘积的合计。

如果使用依时协变量运行 Cox，则只有 DfBeta 和线性预测变量 X*Beta 才可保存。

将模型信息导出到 XML 文件。将参数估计值导出到指定的 XML 格式的文件。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。

Cox 回归：选项

图片 14-5
“Cox 回归：选项”对话框

您可以控制分析和输出的各个方面。

模型统计量。您可以获得模型参数的统计量，包括 exp(B) 的置信区间和估计值的相关性。您可以在每一步或者仅在最后一步请求这些统计量。

步进概率。如果选择了逐步推进方法，可以指定模型的输入或剔除的概率。如果变量的进入 F 的显著性水平小于“进入”值，则输入该变量；如果变量的该显著性水平大于“删除”值，则移去该变量。“进入”值必须小于“删除”值。

最大迭代次数。允许您指定模型的最大迭代次数，用于控制过程求解的时间。

显示基线函数。允许您显示协变量均值下的基线风险函数和累积生存。如果指定了依时协变量，则此显示不可用。

Cox 回归：为状态变量定义事件

输入表示已出现期间终结的值。可以输入单个值、值范围或值列表。只有在状态变量为数值时，“值范围”选项才可用。

COXREG 命令的附加功能

使用命令语法语言还可以：

- 获得将要追踪调查的个案损失作为独立于已审查个案的单独类别的频率表。
- 为偏差、简单和指示对比方法选择除第一个和最后一个以外的参考类别。
- 为多项式对比方法指定类别的不等间距。
- 指定附加迭代标准。
- 控制对缺失值的处理。
- 指定已保存的变量的名称。
- 将输出写入到外部 IBM® SPSS® Statistics 数据文件。
- 在处理过程中将每个拆分文件组的数据保存到一个外部临时文件。这样有助于在运行大型数据集分析时节约内存资源。此功能对于依时协变量不可用。

请参阅命令语法参考以获取完整的语法信息。

章
15

计算依时协变量

在某些情况下，您可能想要计算“Cox 回归”模型，但并不符合成比例的风险假设。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）协变量的值会不同。在这种情况下，您就需要使用扩展的“Cox 回归”模型，该模型允许您指定**依时协变量**。

要想分析这样的模型，您必须首先定义依时协变量。（使用命令语法可以指定多个依时协变量。）使用表示时间的系统变量可以简化此过程。此变量称为 T_。您可以使用此变量通过两种常用方法定义依时协变量：

- 如果您想要针对特定协变量检验成比例的风险假设或者估计允许不成比例的风险的扩展“Cox 回归”模型，则可以通过将依时协变量定义为时间变量 T_ 和有问题的协变量的函数来达到此目的。一个常见的例子就是简单地将时间变量和协变量相乘，不过也可以指定较为复杂的函数。通过检验依时协变量的系数的显著性就可以知道成比例的风险假设是否合理。
- 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**分段依时协变量**，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协变量。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），则可以将依时协变量定义为 (T_ < 1) * BP1 + (T_ >= 1 & T_ < 2) * BP2 + (T_ >= 2 & T_ < 3) * BP3 + (T_ >= 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。

在“计算依时协变量”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“T_COV_ 的表达式”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。得到的变量称为 T_COV_，应该作为协变量包含在“Cox 回归”模型中。

计算依时协变量

▶ 从菜单中选择：

分析 > 生存 > Cox 依时协变量...

图片 15-1
“计算依时协变量”对话框

- 输入依时协变量的表达式。
- 单击**模型**继续“Cox 回归”。

  注意：确保在“Cox 回归”模型中包含新变量 T_COV_ 作为协变量。

带依时协变量的 Cox 回归的附加功能

命令语法语言还允许指定多个依时协变量。其他命令语法功能对所有“Cox 回归”均可用，无论是否带有依时协变量。

请参见命令语法参考以获取完整的语法信息。

附录
A

分类变量编码设计

在许多过程中，可以请求用一组对比变量自动替换分类自变量，该自变量随后将作为一个块输入方程式或从方程式中移除。通常，可以在 CONTRAST 子命令中指定这组对比变量的编码方式。本附录解释并说明 CONTRAST 中所需要的不同对比类型的实际工作方式。

偏差

与总均值的偏差。在矩阵项中，对比具有以下形式：

均值	(1/k	1/k	···	1/k	1/k)
df(1)	(1 - 1/k	- 1/k	···	- 1/k	- 1/k)
df(2)	(- 1/k	1 - 1/k	···	- 1/k	- 1/k)
.		.			
.		.			
df(k - 1)	(- 1/k	- 1/k	···	1 - 1/k	- 1/k)

其中 k 是自变量的类别数量。缺省情况下，省略最后一个类别。例如，一个具有三个类别的自变量的偏移对比为：

(1/3	1/3	1/3)
(2/3	- 1/3	- 1/3)
(- 1/3	2/3	- 1/3)

若要省略除最后一个类别以外的类别，请在 DEVIATION 关键字之后的括号内指定要省略的类别的序号。例如，以下子命令获取第一个和第三个类别的偏差并省略第二个类别：

```
/CONTRAST(FACTOR)=DEVIATION(2)
```

假设因子有三个类别。生成的对比矩阵将是

(1/3	1/3	1/3)
(2/3	- 1/3	- 1/3)
(- 1/3	- 1/3	2/3)

简单散点图

简单对比。将因子的每一级别与上一级别进行比较。一般矩阵格式是

均值	(1/k	1/k	···	1/k	1/k)
df(1)	(1	0	···	0	-1)
df(2)	(0	1	···	0	-1)
.		.			
.		.			
df(k-1)	(0	0	···	1	-1)

其中 k 是自变量类别的数量。例如，具有四个类别的自变量的简单对比如下所示：

(1/4	1/4	1/4	1/4)
(1	0	0	-1)
(0	1	0	-1)
(0	0	1	-1)

若要使用其他类别而不是最后一个类别作为参考类别，请在 SIMPLE 关键字之后的括号中指定参考类别的序号，该序号不必是与该类别相关的值。例如，以下 CONTRAST 子命令获得一个省略了第二个类别的对比矩阵：

```
/CONTRAST(FACTOR) = SIMPLE(2)
```

假设因子有四个类别。生成的对比矩阵将是

(1/4	1/4	1/4	1/4)
(1	-1	0	0)
(0	-1	1	0)
(0	-1	0	1)

Helmert

Helmert 对比。比较自变量的类别与后续类别的均值。一般矩阵格式是

均值	(1/k	1/k	···	1/k	1/k)
df(1)	(1	-1/(k-1)	···	-1/(k-1)	-1/(k-1))
df(2)	(0	1	···	-1/(k-2)	-1/(k-2))
.		.			
.		.			
df(k-2)	(0	0	1	-1/2	-1/2
df(k-1)	(0	0	···	1	-1)

其中 k 是自变量类别的数量。例如，一个有四个类别的自变量具有以下形式的 Helmert 对比矩阵：

```
( 1/4     1/4     1/4     1/4 )
( 1    - 1/3   - 1/3   - 1/3 )
( 0       1    - 1/2   - 1/2 )
( 0       0       1    - 1 )
```

差分

差分或逆 Helmert 对比。比较自变量的类别和该变量的先前类别的均值。一般矩阵格式是

```
   均值      ( 1/k       1/k          1/k    ···     1/k )
  df(1)      ( - 1         1            0    ···       0 )
  df(2)    ( - 1/2     - 1/2          1    ···       0 )
     .                    .
     .                    .
df(k - 1)          (  - 1/(k - 1) - 1/(k - 1)   ···       1 )
          - 1/(k - 1)
```

其中 k 是自变量类别的数量。例如，具有四个类别的自变量的差分对比如下所示：

```
( 1/4     1/4     1/4     1/4 )
( - 1       1       0       0 )
( - 1/2   - 1/2     1       0 )
( - 1/3   - 1/3   - 1/3     1 )
```

多项式

正交多项式对比。第一自由度包含跨所有类别的线性作用；第二自由度包含二次作用；第三自由度包含三次作用；对于更高阶的作用，依此类推。

可以指定由给定的分类变量度量的处理级别之间的间距。相等的间距是省略度规时的缺省值。可以将相等的间距指定为从 1 到 k 的连续整数，其中 k 是类别的数量。如果变量药物有三个类别，则子命令

```
/CONTRAST(DRUG)=POLYNOMIAL
```

等同于

```
/CONTRAST(DRUG)=POLYNOMIAL(1,2,3)
```

然而，并不总是需要相等的间距。例如，假设药物代表某药物分配给三个组的不同剂量。如果第二组的控制剂量是给第一组的剂量的两倍，并且第三组的控制剂量是给第一组的剂量的三倍，则处理类别在间距上是相等的，这种情况下，由连续整数组成的度规比较适用：

```
/CONTRAST(DRUG)=POLYNOMIAL(1,2,3)
```

但是，如果第二组的控制剂量是给第一组的剂量的四倍，并且第三组的控制剂量是给第一组的剂量的七倍，则适用的度规为

```
/CONTRAST(DRUG)=POLYNOMIAL(1,4,7)
```

在每种情况下，对比指定的结果都是药物的第一自由度包含剂量级别的线性作用，而第二自由度包含二次作用。

多项式对比在测试趋势以及调查响应曲面的性质时特别有用。还可以使用多项式对比进行非线性曲线拟合，例如曲线回归。

重复

比较自变量的相邻级别。一般矩阵格式是

均值	(1/k	1/k	1/k	···	1/k	1/k)
df(1)	(1	-1	0	···	0	0)
df(2)	(0	1	-1	···	0	0)
.		.				
.		.				
df(k-1)	(0	0	0	···	1	-1)

其中 k 是自变量类别的数量。例如，具有四个类别的自变量的重复对比如下所示：

(1/4	1/4	1/4	1/4)
(1	-1	0	0)
(0	1	-1	0)
(0	0	1	-1)

这些对比在概要分析以及任何需要不同得分的情况下都很有用。

特殊

一种用户定义的对比。允许以方阵的形式输入特殊对比，该方阵的行和列的数量与给定的自变量的类别数量相等。对于 MANOVA 和 LOGLINEAR，输入的第一行总是均值（或常数）效应，并且代表一组权重，这组权重指示如何根据给定的变量计算其他自变量（如果有）的平均数。通常情况下，该对比为 1 的矢量。

矩阵的其余行包含特殊对比，这些对比指示所需的变量类别之间的比较。一般来说，正交对比是最有用的。正交对比在统计上相互独立并且无冗余。在下列情况下，对比是正交对比：

- 每行的对比系数和都为 0。
- 每对非联合行所对应的系数的积的和也为 0。

例如，假设处理有四个级别并且要在各个处理级别之间进行比较。以下是一个对应的特殊对比

(1	1	1	1)	均值计算的权重
(3	-1	-1	-1)	将第 1 级别与第 2 到第 4 级别进行比较
(0	2	-1	-1)	将第 2 级别与第 3 和第 4 级别比较
(0	0	1	-1)	比较第 3 级别与第 4 级别

该对比通过以下命令的 CONTRAST 子命令指定：MANOVA、LOGISTIC REGRESSION 和 COXREG：

```
/CONTRAST(TREATMNT)=SPECIAL( 1  1  1  1
                     3 -1 -1 -1
                     0  2 -1 -1
                     0  0  1 -1 )
```

对于 LOGLINEAR，需要指定：

```
/CONTRAST(TREATMNT)=BASIS SPECIAL( 1  1  1  1
                          3 -1 -1 -1
                          0  2 -1 -1
                          0  0  1 -1 )
```

除均值行之外的每行的和都为 0。每对非联合行的积的和也为 0：

行 2 和 3：	(3)(0) + (-1)(2) + (-1)(-1) + (-1)(-1) = 0
行 2 和 4：	(3)(0) + (-1)(0) + (-1)(1) + (-1)(-1) = 0
行 3 和 4：	(0)(0) + (2)(0) + (-1)(1) + (-1)(-1) = 0

特殊对比不需要是正交的。但是，特殊对比不能是每个对比的线性组合。如果是，则过程会报告线性相关性并终止处理。Helmert、差分和多项式对比都是正交对比。

指示符

指示变量编码。也称为哑元编码，在 LOGLINEAR 或 MANOVA 中不可用。新编码变量的数量是 k - 1。参考类别中的个案将所有 k - 1 个变量都编码为 0。第 i 个类别中的个案将除第 i 个变量（编码为 1）之外的所有指示变量都编码为 0。

附录

B

协方差结构

本节提供有关协方差结构的额外信息。

前因:一阶。此协方差结构在相邻元素之间具有异质方差和异质相关性。非相邻元素之间的相关性是位于所涉及元素之间的元素间相关性的乘积。

$$\begin{bmatrix} \sigma_1^2 & \sigma_2\sigma_1\rho_1 & \sigma_3\sigma_1\rho_1\rho_2 & \sigma_4\sigma_1\rho_1\rho_2\rho_3 \\ \sigma_2\sigma_1\rho_1 & \sigma_2^2 & \sigma_3\sigma_2\rho_2 & \sigma_4\sigma_2\rho_2\rho_3 \\ \sigma_3\sigma_1\rho_1\rho_2 & \sigma_3\sigma_2\rho_2 & \sigma_3^2 & \sigma_4\sigma_3\rho_3 \\ \sigma_4\sigma_1\rho_1\rho_2\rho_3 & \sigma_4\sigma_2\rho_2\rho_3 & \sigma_4\sigma_3\rho_3 & \sigma_4^2 \end{bmatrix}$$

AR(1)。这是具有同质方差的一阶自回归结构。任意两个元素之间的相关性对于相邻元素为 rho，对于由第三个元素分隔的元素为 ρ^2，依此类推。ρ 受到约束，以使 $-1<\rho<1$。

$$\sigma^2 \begin{bmatrix} 1 & \rho & \rho^2 & \rho^3 \\ \rho & 1 & \rho & \rho^2 \\ \rho^2 & \rho & 1 & \rho \\ \rho^3 & \rho^2 & \rho & 1 \end{bmatrix}$$

AR(1):异质。这是具有异质方差的一阶自回归结构 。任意两个元素之间的相关性对于相邻元素为 ρ，对于由第三个元素分隔的两个元素为 ρ^2，依此类推。ρ 约束在 -1 和 1 之间。

$$\begin{bmatrix} \sigma_1^2 & \sigma_2\sigma_1\rho & \sigma_3\sigma_1\rho^2 & \sigma_4\sigma_1\rho^3 \\ \sigma_2\sigma_1\rho & \sigma_2^2 & \sigma_3\sigma_2\rho & \sigma_4\sigma_2\rho^2 \\ \sigma_3\sigma_1\rho^2 & \sigma_3\sigma_2\rho & \sigma_3^2 & \sigma_4\sigma_3\rho \\ \sigma_4\sigma_1\rho^3 & \sigma_4\sigma_2\rho^2 & \sigma_4\sigma_3\rho & \sigma_4^2 \end{bmatrix}$$

ARMA(1,1)。这是一阶自回归移动平均结构。它具有同质方差。两个元素之间的相关性对于相邻元素为 ϕ*ρ，对于由第三个元素分隔的元素为 $\phi*(\rho^2)$，依此类推。ρ 和 ϕ 分别为自回归和移动平均参数，它们的值约束在 -1 和 1 之间（含边界）。

$$\sigma^2 \begin{bmatrix} 1 & \phi\rho & \phi\rho^2 & \phi\rho^3 \\ \phi\rho & 1 & \phi\rho & \phi\rho^2 \\ \phi\rho^2 & \phi\rho & 1 & \phi\rho \\ \phi\rho^3 & \phi\rho^2 & \phi\rho & 1 \end{bmatrix}$$

复合对称。此结构具有常数方差和常数协方差。

$$\begin{bmatrix} \sigma^2+\sigma_1^2 & \sigma_1 & \sigma_1 & \sigma_1 \\ \sigma_1 & \sigma^2+\sigma_1^2 & \sigma_1 & \sigma_1 \\ \sigma_1 & \sigma_1 & \sigma^2+\sigma_1^2 & \sigma_1 \\ \sigma_1 & \sigma_1 & \sigma_1 & \sigma^2+\sigma_1^2 \end{bmatrix}$$

复合对称：相关性度规。此协方差结构的元素之间具有同质方差和同质相关性。

$$\sigma^2 \begin{bmatrix} 1 & \rho & \rho & \rho \\ \rho & 1 & \rho & \rho \\ \rho & \rho & 1 & \rho \\ \rho & \rho & \rho & 1 \end{bmatrix}$$

复合对称：异质。此协方差结构在元素之间具有异质方差和常数相关性。

$$\begin{bmatrix} \sigma_1^2 & \sigma_2\sigma_1\rho & \sigma_3\sigma_1\rho & \sigma_4\sigma_1\rho \\ \sigma_2\sigma_1\rho & \sigma_2^2 & \sigma_3\sigma_2\rho & \sigma_4\sigma_2\rho \\ \sigma_3\sigma_1\rho & \sigma_3\sigma_2\rho & \sigma_3^2 & \sigma_4\sigma_3\rho \\ \sigma_4\sigma_1\rho & \sigma_4\sigma_2\rho & \sigma_4\sigma_3\rho & \sigma_4^2 \end{bmatrix}$$

对角线。此协方差结构在元素之间具有异质方差和零相关性。

$$\begin{bmatrix} \sigma_1^2 & 0 & 0 & 0 \\ 0 & \sigma_2^2 & 0 & 0 \\ 0 & 0 & \sigma_3^2 & 0 \\ 0 & 0 & 0 & \sigma_4^2 \end{bmatrix}$$

因子分析：一阶。此协方差结构 具有异质方差，该方差由元素间的一个异质项和一个同质项组成。任意两个元素之间的协方差是它们的异质方差项乘积的平方根。

$$\begin{bmatrix} \lambda_1^2+d & \lambda_2\lambda_1 & \lambda_3\lambda_1 & \lambda_4\lambda_1 \\ \lambda_2\lambda_1 & \lambda_2^2+d & \lambda_3\lambda_2 & \lambda_4\lambda_2 \\ \lambda_3\lambda_1 & \lambda_3\lambda_2 & \lambda_3^2+d & \lambda_4\lambda_3 \\ \lambda_4\lambda_1 & \lambda_4\lambda_2 & \lambda_4\lambda_3 & \lambda_4^2+d \end{bmatrix}$$

因子分析：一阶，异质。此协方差结构具有异质方差，这些方差由元素间的两个异质项组成。任意两个元素之间的协方差 是它们的第一个异质方差项乘积的平方根。

$$\begin{bmatrix} \lambda_1^2+d_1 & \lambda_2\lambda_1 & \lambda_3\lambda_1 & \lambda_4\lambda_1 \\ \lambda_2\lambda_1 & \lambda_2^2+d_2 & \lambda_3\lambda_2 & \lambda_4\lambda_2 \\ \lambda_3\lambda_1 & \lambda_3\lambda_2 & \lambda_3^2+d_3 & \lambda_4\lambda_3 \\ \lambda_4\lambda_1 & \lambda_4\lambda_2 & \lambda_4\lambda_3 & \lambda_4^2+d_4 \end{bmatrix}$$

Huynh-Feldt。这是一个“圆形”矩阵，其中任意两个元素之间的协方差等于它们的方差平均值减去一个常数。 方差和协方差都不是常数。

$$\begin{bmatrix} \sigma_1^2 & \frac{\sigma_1^2+\sigma_2^2}{2}-\lambda & \frac{\sigma_1^2+\sigma_3^2}{2}-\lambda & \frac{\sigma_1^2+\sigma_4^2}{2}-\lambda \\ \frac{\sigma_1^2+\sigma_2^2}{2}-\lambda & \sigma_2^2 & \frac{\sigma_2^2+\sigma_3^2}{2}-\lambda & \frac{\sigma_2^2+\sigma_4^2}{2}-\lambda \\ \frac{\sigma_1^2+\sigma_3^2}{2}-\lambda & \frac{\sigma_2^2+\sigma_3^2}{2}-\lambda & \sigma_3^2 & \frac{\sigma_3^2+\sigma_4^2}{2}-\lambda \\ \frac{\sigma_1^2+\sigma_4^2}{2}-\lambda & \frac{\sigma_2^2+\sigma_4^2}{2}-\lambda & \frac{\sigma_3^2+\sigma_4^2}{2}-\lambda & \sigma_4^2 \end{bmatrix}$$

已标度的恒等。此结构具有常数方差。假设任意两个元素之间没有相关性。

$$\sigma^2 \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}$$

Toeplitz。 此协方差结构的元素之间具有同质方差和 异质相关性。不同相邻元素对的相邻元素之间的相关性是同质的。被第三个元素分隔开的元素之间的相关性又是同质的，依此类推。

$$\sigma^2 \begin{bmatrix} 1 & \rho_1 & \rho_2 & \rho_3 \\ \rho_1 & 1 & \rho_1 & \rho_2 \\ \rho_2 & \rho_1 & 1 & \rho_1 \\ \rho_3 & \rho_2 & \rho_1 & 1 \end{bmatrix}$$

Toeplitz：异质。此协方差结构在元素之间具有异质方差和异质相关性。不同相邻元素对的相邻元素之间的相关性是同质的。被第三个元素分隔开的元素之间的相关性又是同质的，依此类推。

$$\begin{bmatrix} \sigma_1^2 & \sigma_2\sigma_1\rho_1 & \sigma_3\sigma_1\rho_2 & \sigma_4\sigma_1\rho_3 \\ \sigma_2\sigma_1\rho_1 & \sigma_2^2 & \sigma_3\sigma_2\rho_1 & \sigma_4\sigma_2\rho_2 \\ \sigma_3\sigma_1\rho_2 & \sigma_3\sigma_2\rho_1 & \sigma_3^2 & \sigma_4\sigma_3\rho_1 \\ \sigma_4\sigma_1\rho_3 & \sigma_4\sigma_2\rho_2 & \sigma_4\sigma_3\rho_1 & \sigma_4^2 \end{bmatrix}$$

未结构化。这是一个非常一般的协方差矩阵。

$$\begin{bmatrix} \sigma_1^2 & \sigma_{21} & \sigma_{31} & \sigma_{41} \\ \sigma_{21} & \sigma_2^2 & \sigma_{32} & \sigma_{42} \\ \sigma_{31} & \sigma_{32} & \sigma_3^2 & \sigma_{43} \\ \sigma_{41} & \sigma_{42} & \sigma_{43} & \sigma_4^2 \end{bmatrix}$$

未结构化：相关性度规。此协方差结构具有异质方差和异质相关性。

$$\begin{bmatrix} \sigma_1^2 & \sigma_2\sigma_1\rho_{21} & \sigma_3\sigma_1\rho_{31} & \sigma_4\sigma_1\rho_{41} \\ \sigma_2\sigma_1\rho_{21} & \sigma_2^2 & \sigma_3\sigma_2\rho_{32} & \sigma_4\sigma_2\rho_{42} \\ \sigma_3\sigma_1\rho_{31} & \sigma_3\sigma_2\rho_{32} & \sigma_3^2 & \sigma_4\sigma_3\rho_{43} \\ \sigma_4\sigma_1\rho_{41} & \sigma_4\sigma_2\rho_{42} & \sigma_4\sigma_3\rho_{43} & \sigma_4^2 \end{bmatrix}$$

方差成分。此结构为每个指定的随机效应分配一个已标度的恒等 (ID) 结构。

附录
C

Notices

Patent No. 7,023,453

The following paragraph does not apply to the United Kingdom or any other country where such provisions are inconsistent with local law: SPSS INC., AN IBM COMPANY, PROVIDES THIS PUBLICATION "AS IS" WITHOUT WARRANTY OF ANY KIND, EITHER EXPRESS OR IMPLIED, INCLUDING, BUT NOT LIMITED TO, THE IMPLIED WARRANTIES OF NON-INFRINGEMENT, MERCHANTABILITY OR FITNESS FOR A PARTICULAR PURPOSE. Some states do not allow disclaimer of express or implied warranties in certain transactions, therefore, this statement may not apply to you.

This information could include technical inaccuracies or typographical errors. Changes are periodically made to the information herein; these changes will be incorporated in new editions of the publication. SPSS Inc. may make improvements and/or changes in the product(s) and/or the program(s) described in this publication at any time without notice.

Any references in this information to non-SPSS and non-IBM Web sites are provided for convenience only and do not in any manner serve as an endorsement of those Web sites. The materials at those Web sites are not part of the materials for this SPSS Inc. product and use of those Web sites is at your own risk.

When you send information to IBM or SPSS, you grant IBM and SPSS a nonexclusive right to use or distribute the information in any way it believes appropriate without incurring any obligation to you.

Information concerning non-SPSS products was obtained from the suppliers of those products, their published announcements or other publicly available sources. SPSS has not tested those products and cannot confirm the accuracy of performance, compatibility or any other claims related to non-SPSS products. Questions on the capabilities of non-SPSS products should be addressed to the suppliers of those products.

This information contains examples of data and reports used in daily business operations. To illustrate them as completely as possible, the examples include the names of individuals, companies, brands, and products. All of these names are fictitious and any similarity to the names and addresses used by an actual business enterprise is entirely coincidental.

COPYRIGHT LICENSE:

This information contains sample application programs in source language, which illustrate programming techniques on various operating platforms. You may copy, modify, and distribute these sample programs in any form without payment to SPSS Inc., for the purposes of developing, using, marketing or distributing application programs conforming to the application programming interface for the operating platform for which the sample programs are written. These examples have not been thoroughly tested under all conditions. SPSS Inc., therefore, cannot guarantee or imply reliability, serviceability, or function of these programs. The sample programs are provided “AS IS”, without warranty of any kind. SPSS Inc. shall not be liable for any damages arising out of your use of the sample programs.

Trademarks

IBM, the IBM logo, and ibm.com are trademarks of IBM Corporation, registered in many jurisdictions worldwide. A current list of IBM trademarks is available on the Web at http://www.ibm.com/legal/copytrade.shmtl.

SPSS is a trademark of SPSS Inc., an IBM Company, registered in many jurisdictions worldwide.

Adobe, the Adobe logo, PostScript, and the PostScript logo are either registered trademarks or trademarks of Adobe Systems Incorporated in the United States, and/or other countries.

Intel, Intel logo, Intel Inside, Intel Inside logo, Intel Centrino, Intel Centrino logo, Celeron, Intel Xeon, Intel SpeedStep, Itanium, and Pentium are trademarks or registered trademarks of Intel Corporation or its subsidiaries in the United States and other countries.

Linux is a registered trademark of Linus Torvalds in the United States, other countries, or both.

Microsoft, Windows, Windows NT, and the Windows logo are trademarks of Microsoft Corporation in the United States, other countries, or both.

UNIX is a registered trademark of The Open Group in the United States and other countries.

Java and all Java-based trademarks and logos are trademarks of Sun Microsystems, Inc. in the United States, other countries, or both.

This product uses WinWrap Basic, Copyright 1993-2007, Polar Engineering and Consulting, http://www.winwrap.com.

Other product and service names might be trademarks of IBM, SPSS, or other companies.

Adobe product screenshot(s) reprinted with permission from Adobe Systems Incorporated.

Microsoft product screenshot(s) reprinted with permission from Microsoft Corporation.

索引