

*IBM SPSS Advanced Statistics 30*

**IBM**

## 注

使用本信息及其支持的产品前，请先阅读 [第 107 页的『声明』](#) 中的信息。

## 产品信息

本版本适用于 30 版本 0、IBM® SPSS Statistics 的修改 0 以及所有后续版本和修改，除非新版本另有说明。

© Copyright International Business Machines Corporation .

# 内容

<b>第 1 章 高级统计</b> .....	<b>1</b>
Advanced Statistics 简介.....	1
GLM 多变量分析.....	1
GLM 多变量: 模型.....	2
GLM 多变量: 对比.....	3
GLM 多变量: 概要图.....	3
GLM 多变量: 事后比较.....	4
GLM 估计边际均值.....	5
GLM 保存.....	5
GLM 多变量: 选项.....	5
GLM 命令的附加功能.....	6
GLM 重复测量.....	7
GLM 重复测量定义因子.....	8
GLM 重复测量: 模型.....	9
GLM 重复测量: 对比.....	10
GLM 重复测量: 概要图.....	10
GLM 重复测量: 事后比较.....	11
GLM 估计边际均值.....	11
GLM 重复测量: 保存.....	12
GLM 重复测量: 选项.....	12
GLM 命令的附加功能.....	13
方差成分分析.....	13
方差成分: 模型.....	14
方差成分: 选项.....	15
方差成分: 保存到新文件.....	15
VARCOMP 命令的附加功能.....	16
线性混合模型.....	16
线性混合模型: 主体与重复.....	17
线性混合模型: 固定效应.....	18
线性混合模型: 随机效果.....	19
线性混合模型: 估算.....	20
线性混合模型: 统计.....	20
线性混合模型: EM 均值.....	21
线性混合模型: 保存.....	21
线性混合模型-导出.....	21
MIXED 命令的附加功能.....	22
广义线性模型.....	22
广义线性模型响应.....	24
广义线性模型: 预测变量.....	24
广义线性模型: 模型.....	25
广义线性模型: 估计.....	25
广义线性模型: 统计.....	26
广义线性模型: EM 均值.....	27
广义线性模型: 保存.....	27
广义线性模型: 导出.....	28
GENLIN 命令的附加功能.....	29
广义估算方程.....	29
广义估计方程: 模型类型.....	30
广义估计方程: 响应.....	32
广义估计方程: 预测变量.....	32
广义估计方程: 模型.....	33

广义估计方程: 估计.....	33
广义估计方程: 统计.....	35
广义估计方程: EM 均值.....	35
广义估计方程: 保存.....	36
广义估计方程: 导出.....	36
GENLIN 命令的附加功能.....	37
广义线性混合模型.....	37
获取广义线性混合模型.....	39
目标.....	39
固定效应.....	40
随机效应.....	41
权重和偏移量.....	42
常规构建选项.....	43
估计.....	43
估计均值.....	43
保存.....	44
导出.....	44
模型视图.....	45
模型选择对数线性分析.....	47
对数线性分析: 定义范围.....	48
对数线性分析: 模型.....	48
模型选择对数线性分析: 选项.....	49
HILOGLINEAR 命令的附加功能.....	49
一般对数线性分析.....	49
一般对数线性分析模型.....	50
一般对数线性分析: 选项.....	50
一般对数线性分析: 保存.....	51
GENLOG 命令的附加功能.....	51
分对数的对数线性分析.....	51
Logit 对数线性分析: 模型.....	52
Logit 对数线性分析: 选项.....	53
Logit 对数线性分析: 保存.....	53
GENLOG 命令的附加功能.....	53
寿命表.....	53
寿命表: 定义范围.....	54
寿命表: 选项.....	54
SURVIVAL 命令的附加功能.....	55
参数加速故障时间模型.....	55
参数加速故障时间模型: 条件.....	56
参数加速故障时间模型: 模型.....	56
参数加速故障时间模型: 估算.....	57
参数加速故障时间模型: 打印.....	58
参数加速故障时间模型: 预测.....	58
参数加速故障时间模型: 图.....	59
参数加速故障时间模型: 导出.....	59
生存 AFT 定义状态变量的事件.....	59
参数加速故障时间模型: 选择类别.....	60
参数化共享脆弱模型.....	60
参数化共享脆弱模型: 条件.....	61
参数化共享脆弱模型: 模型.....	62
参数化共享脆弱模型: 估算.....	62
参数化共享脆弱模型: 打印.....	64
参数共享脆弱模型: 预测.....	64
参数化共享脆弱模型: 图.....	65
参数化共享脆弱模型: 导出.....	65
参数化共享脆弱模型: 定义事件.....	65
参数化共享脆弱模型-示例.....	66
参数化共享脆弱模型-经常性数据的案例研究.....	68

Kaplan-Meier 生存分析.....	69
Kaplan-Meier: 为状态变量定义事件.....	70
Kaplan-Meier: 比较因子级别.....	70
Kaplan-Meier: 保存新变量.....	70
Kaplan-Meier: 选项.....	71
KM 命令的附加功能.....	71
Cox 回归分析.....	71
Cox 回归: 定义分类变量.....	72
Cox 回归: 图.....	72
Cox 回归: 保存新变量.....	72
Cox 回归: 选项.....	73
Cox 回归: 为状态变量定义事件.....	73
COXREG 命令的附加功能.....	73
计算依时协变量.....	74
计算依时协变量.....	74
分类变量编码方案.....	74
偏差.....	74
简单.....	75
赫尔默特.....	75
差分.....	75
多项式.....	76
重复.....	76
特殊.....	76
指示符.....	77
协方差结构.....	77
Storing transaction logs in a database for high availability.....	81
贝叶斯单一样本推论: 正态.....	82
贝叶斯单一样本推论: 二项式.....	85
贝叶斯单一样本推论: 泊松.....	87
贝叶斯相关样本推论: 正态.....	88
贝叶斯独立样本推论.....	89
关于 Pearson 相关性的贝叶斯推论.....	92
关于线性回归模型的贝叶斯推论.....	94
贝叶斯单因素 ANOVA.....	98
贝叶斯对数线性模型.....	100
贝叶斯单向重复测量 ANOVA 模型.....	102
内核岭回归.....	104
内核参数.....	105
核岭回归: 选项.....	105
<b>声明.....</b>	<b>107</b>
商标.....	108
<b>索引.....</b>	<b>109</b>



# 第 1 章 高级统计

SPSS Statistics Standard Edition 或“高级统计”选项中包含以下高级统计功能。

## Advanced Statistics 简介

"高级统计信息" 选项包含提供比通过 "统计信息库" 选项可用的更高级建模选项的过程。

- “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”允许重复测量多个因变量。
- “方差成分分析”是将因变量的可变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可变性。因此，线性混合模型提供了不仅能够就数据的平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，以允许重复测量。
- “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。

## 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平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格包含相同的案例数，那么设计是平衡的。在多变量模型中，由于模型中的影响而产生的平方和和误差平方和采用矩阵形式，而不是单变量分析中的标量形式。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，可以使用事后检验评估指定平均值之间的差值。估计的边际均值提供模型中单元格的预测平均值的估计值，这些均值的剖面图（交互图）使您可以轻松地将某些关系可视化。单独为每个因变量执行两两多重比较检验。

残差、预测值、库克距离和杠杆值可以保存为数据文件中的新变量，以检查假设。还可以使用残差 SSCP 矩阵（残差的平方和和叉积的方阵）、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，这是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，用于为加权最小二乘 (WLS) 分析赋予观测值不同的权重，或许是为了补偿不同的测量精度。

**示例。** 某塑料制造商要测量塑料膜的三种属性：耐撕裂性、光泽和不透明度。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。** 两两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's  $b$ 、Duncan、Hochberg's GT2、Gabriel、Waller Duncan  $t$  检验、Dunnnett（单侧和双侧）、Tamhane's T2、Dunnnett's T3、Games-Howell 和 Dunnnett's C。描述统计：所有单元格中所有因变量的观

察平均值、标准差和计数；Levene 的方差同质性检验；对因变量协方差矩阵的同质性 Box 的 *M* 检验以及 Bartlett 的球形度检验。

图。分布-水平图、残差图以及概要图（交互）。

### GLM 多变量数据注意事项

**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，所有单元格的方差-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，您可以使用方差同质性检验（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

1. 从菜单中选择：

**分析 > 一般线性模型 > 多变量...**

2. 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

**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。对于没有缺失单元格的平衡或非平衡模型，类型 III 平方和法最常用。

**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。

## 构建项和定制项

**构建项：**

    当您想要为一组选定因子和协变量的所有组合包含特定类型（例如主效应）的非嵌套项时，使用该选项。

**构建定制项**

    如果要包含嵌套项，或者想要按变量显式构建任何项，请使用该选项。构建嵌套项包括以下步骤：

## 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层解构法。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：据任何不包含该效应的其他效应，以及与包含该效应的任何效应的正交（如果存在）进行调整的平方和。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，平方和对于单元格频率就保持不变。因此，我们常认为此类平方和对于不带缺失单元格的平衡模型有用。在不带缺失单元格的因子设计中，此方法等同于 Yates 加权均方方法。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。
- 任何不带空白单元格的平衡或非平衡模型。

**类型 IV。** 此方法针对存在缺失单元格的情况设计。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。
- 任何带有空白单元格的平衡或非平衡模型。

## GLM 多变量：对比

“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  $\mathbf{LBM} = \mathbf{0}$ ，其中  $\mathbf{L}$  是对比系数矩阵， $\mathbf{M}$  是恒等矩阵，该矩阵的维数等于因变量的个数， $\mathbf{B}$  是参数向量。当指定对比之后，创建一个  $\mathbf{L}$  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  $\mathbf{L}$  矩阵可以估计。

除了使用  $F$  统计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）的平均值与所有水平的平均值（总平均值）进行比较。因子的水平可以为任何顺序。

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

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

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

**重复。** 将每个水平的平均值（最后一个水平除外）与后一个水平的平均值进行比较。

**多项式。** 比较线性效应、二次效应、三次效应等等。第一自由度包含跨所有类别的线性效应；第二自由度包含二次效应，依此类推。这些对比常常用来估计多项式趋势。

## GLM 多变量：概要图

轮廓图（交互图）适合用来比较模型中的边际均值。概要图是一个线图，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）。第二个因子的级别可用来绘制单独的线条。第三个因子中的每个级别可用于创建单独的图。所有因子都可用于图。为每个因变量创建概要图。

单因子的概要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。

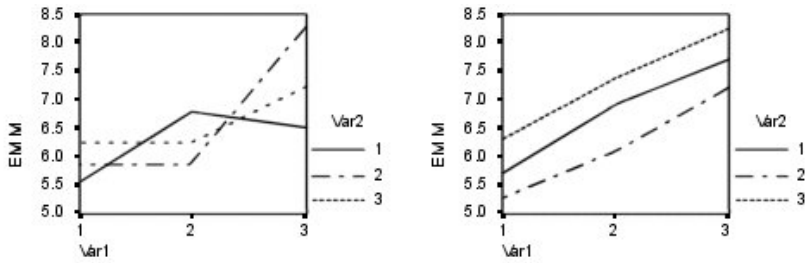


图 1: 不平行图 (左) 和平行图 (右)

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：事后比较

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

**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，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。当检验大量平均值对时，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 Tukey's 真实类显著性差异检验**，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。当单元格大小变化过大时，Gabriel 检验可能会变得随意。

**Dunnnett 的成对多重比较  $t$  检验**将一组处理与单个控制平均值进行比较。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尾检验。要检验因子的任何水平（控制类别除外）的平均值是否不等于控制类别的平均值，请使用双侧检验。要检验因子的任何水平的平均值是否小于控制类别的平均值，请选择  $<$  控制。同样，要检验因子的任何水平的平均值是否大于控制类别的平均值，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平均值是否相等。如果不是所有的平均值均相等，则检验一部分平均值的等同性。**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格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）、**Dunnnett 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 **Dunnnett'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** 是排列组平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平均值。

**Scheffé 检验**的显著性水平可允许要检验的组平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，需要平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 T2 和 T3、Dunnnett's C 以及 Dunnnett T3 提供成对比较。为 S-N-K、Tukey  $b$ 、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's 的检验既是多重比较检验，同时也是多范围检验。

## GLM 估计边际均值

选择您需要的单元格中的总体边际平均值估计的因子和交互作用。为协变量（如果存在）调整这些平均值。

### 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平均值未修正的成对比较。只有在“显示以下项的平均值”列表中选择了主效应的情况下，此项才可用。

### 比较简单主效应

只要目标列表包含一个或多个乘积或交互效应（例如， $A \times B$  和  $A \times B \times C$ ），就会启用该设置。该设置支持指定简单主效应之间的比较，其中，这些效应是嵌套在其他因子的级别内的主效应。

### 置信区间调整

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**比较主效应**和/或**比较简单主效应**的情况下才可用。

## 指定估计边际均值

1. 从 > 分析 > 一般线性模型下选择一种过程。
2. 在主对话框中，单击 **EM 平均值**。

## GLM 保存

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。模型预测的因变量的值。
- 加权。加权非标准化预测值。仅在之前已选择了 WLS 变量的情况下可用。
- 标准误差。对于自变量具有相同值的观测值所对应的因变量的均值的标准差的估计。

**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。这是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测量。较大的 Cook 距离表明从回归统计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。每个观测值对模型拟合度的相对影响。

**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，那么提供加权的未标准化残差。

- 未标准化。观测值和模型预测值之间的差值。
- 加权。加权非标准化残差。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值。标准化残差也称为 Pearson 残差，它的平均值为 0，标准差为 1。
- Student 化。残差除以其标准差的估计值，该估计值因个案而异，具体取决于自变量上每个个案的值与自变量均数的距离。有时称为内部学生化残差。
- 剔除。当某个观测值从回归系数的计算中排除时，该观测值的残差。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 IBM 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、一行参数估计的标准误差、一行与参数估计值对应的  $t$  统计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。在选中异方差性一致性统计（仅适用于单变量模型）时，将使用一个稳健估计计算协方差矩阵，标准误差行显示稳健标准误差，而显著性值反映稳固误差。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

此对话框中有一些可选统计。统计是使用固定效应模型计算的。

**输出。** 选择**描述统计**以生成所有单元格中的所有因变量的观察到的平均值、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  $\eta^2$  方值。 $\eta^2$  方统计描述总可变量中可归因于某个因子的部分。当基于

观察值设置备用假设时，选择**观察势**可获取检验的势。选择**参数估计**可为每个检验生成参数估计值、标准误差、*t*检验、置信区间和检验的观察势。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。

**同质性检验**为跨主体间因子所有水平组合的每个因变量生成 Levene 的方差同质性检验（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，那么禁用此项。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都是一般可估计函数的线性组合。

## 显示

### 描述性统计

生成所有单元格中的所有因变量的观察平均值、标准差和计数。

### 效应大小估算

给出每个效应和每个参数估计的部分 eta 方值。eta 方统计描述总可变性中可归因于某个因子的部分。

### 实测幂

基于实测值来设置备择假设时，获取该检验的功效。

### 参数估算值

每个检验生成参数估计值、标准误差、*t*检验、置信区间和检验的观察势。

### SSCP 矩阵

显示假设和误差 SSCP 矩阵。

### 残差 SSCP 矩阵

显示假设和误差残差 SSCP 矩阵。

### 转换矩阵

显示残差协方差矩阵的巴特利特球形度检验。

### 同质性检验

生成跨主体间因子的所有水平组合的每个因变量的方差同质性的莱文检验（仅针对主体间因子）。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

对于检查数据相关假设以调查方差相等的假设很有用。如果不存在任何因子，那么禁用此项。

### 残差图

为每个因变量生成观察-预测-标准化残差图。此图对于调查方差相等的假设很有用。

### 失拟

检查因变量和自变量之间的关系能否由模型充分地描述。

### 一般可估函数

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都是一般可估计函数的线性组合。

### 显著性水平

您可能想要调整事后检验中使用的显著性水平和用于构建置信区间的置信水平。指定值还用于计算检验的检验效能。如果指定了显著性水平，那么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）。
- 为事后比较指定误差项（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）。
- 构造相关性矩阵数据文件（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）。

请参阅命令语法参考以获取完整的语法信息。

## 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。

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格包含相同的案例数，那么设计是平衡的。在多变量模型中，由于模型中的影响而产生的平方和和误差平方和采用矩阵形式，而不是单变量分析中的标量形式。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，可以使用事后检验评估指定平均值之间的差值。估计的边际均值提供模型中单元格的预测平均值的估计值，这些均值的剖面图（交互图）使您可以轻松地将某些关系可视化。

残差、预测值、库克距离和杠杆值可以保存为数据文件中的新变量，以检查假设。还可以使用残差 **SSCP** 矩阵（残差的平方和和叉积的方阵）、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，这是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，用于为加权最小二乘 (**WLS**) 分析赋予观测值不同的权重，或许是为了补偿不同的测量精度。

**示例。** 根据学生在焦虑等级测试中的分数，将 12 名学生分配到高焦虑或低焦虑组。焦虑等级被称为被试间因素，因为它将被试分成不同的组。学生每人接受四次学习任务的试验，并记录每次试验的错误数。每个试验的错误记录在单独的变量中，并且四个试验的四个级别定义了一个受试者内因素（试验）。试验效果很明显，而试验与焦虑的相互作用不明显。

**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。** 事后范围检验和多重比较（针对受试者之间的因素）：最小显著性差异、Bonferroni、Sidak、Scheffé、Ryan-Einot-Gabriel-Welsch multiple *F*、Ryan-Einot-Gabriel-Welsch（多范围）、Student-Newman-Keuls、Tukey 的诚实显著性差异、Tukey 的 *b* 检验、Duncan 检验、Hochberg 的 GT2 检验、Gabriel 检验、Waller Duncan *t* 检验、Dunnnett（单边和双边）、Tamhane's T2、Dunnnett's T3、Games-Howell 和 Dunnnett's *C*。描述性统计：所有单元格中所有因变量的观察均值、标准差和计数；方差齐性的 Levene 检验；Box 的 *M* 和 Mauchly 的球形检验。

**图。** 分布-水平图、残差图以及概要图（交互）。

GLM 重复测量数据注意事项

**数据。** 因变量应是定量的。主体间因子将样本划分为离散子组，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。为级别数等于重复次数的组定义主体内因子。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，那么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（也称为拆分图或混合模型方法）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格内应该都相同。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，此检验的表现一般。对于较大的样本，此检验的效果可能很显著，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，那么可采用球形度假设。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。“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格内应该都相同。要检验方差-协方差矩阵是否在所有单元格内都相同，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，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，那么可以使用“成对样本 *T* 检验”过程。

获取 GLM 重复测量

1. 从菜单中选择：

**分析 > 一般线性模型 > 重复测量...**

2. 主体内部因子名称及其级别数。

3. 单击**添加**。

4. 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：

5. 键入测量名称。

6. 单击**添加**。

    在定义完所有因子和测量后：

7. 单击**定义**。

8. 在列表上选择一个与主体内部因子（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。

要更改主体内部因子，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。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

**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。对于没有缺失单元格的平衡或非平衡主体间模型，类型 III 平方和法最常用。

## 构建项和定制项

**构建项：**

当您想要为一组选定因子和协变量的所有组合包含特定类型（例如主效应）的非嵌套项时，使用该选项。

**构建定制项**

如果要包含嵌套项，或者想要按变量显式构建任何项，请使用该选项。构建嵌套项包括以下步骤：

## 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层解构法。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：据任何不包含该效应的其他效应，以及与包含该效应的任何效应的正交（如果存在）进行调整的平方和。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，平方和对于单元格频率就保持不变。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。在不带缺失单元格的因子设计中，此方法等同于 Yates 加权均方方法。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。
- 任何不带空白单元格的平衡或非平衡模型。

**类型 IV。** 此方法针对存在缺失单元格的情况设计。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。
- 任何带有空白单元格的平衡或非平衡模型。

## GLM 重复测量：对比

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  $L\mathbf{B} = \mathbf{0}$ ，其中  $L$  是对比系数矩阵， $B$  是参数向量， $M$  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。** 将每个水平的平均值（参考类别除外）与所有水平的平均值（总平均值）进行比较。因子的水平可以为任何顺序。

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

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

**Helmert。** 将因子的每个水平的平均值（最后一个除外）与后续水平的平均值进行比较。

**重复。** 将每个水平的平均值（最后一个水平除外）与后一个水平的平均值进行比较。

**多项式。** 比较线性效应、二次效应和三次效应等。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：概要图

轮廓图（交互图）适合用来比较模型中的边际均值。概要图是一个线图，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）。第二个因子的级别可用来绘制单独的线条。第三个因子中的每个级别可用于创建单独的图。所有因子都可用于图。为每个因变量创建概要图。主体间因子和主体内因子都可用在概要图中。

单因子的概要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。

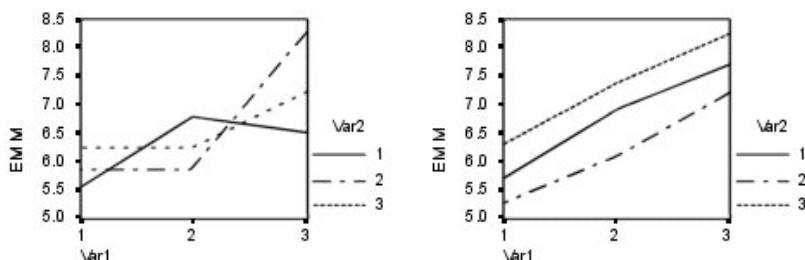


图 2: 不平行图 (左) 和平行图 (右)

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：事后比较

**事后多重比较检验。**一旦确定平均值间存在差值，两两范围检验和成对多重比较就可以确定哪些平均值存在差值了。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平均值执行事后多重比较检验。

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，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。当检验大量平均值对时，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** Tukey's 真实类显著性差异检验，但使用了 Student 化的最大值模数。通常 Tukey 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。当单元格大小变化过大时，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平均值进行比较。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尾检验。要检验因子的任何水平（控制类别除外）的平均值是否不等于控制类别的平均值，请使用双侧检验。要检验因子的任何水平的平均值是否小于控制类别的平均值，请选择 **< 控制**。同样，要检验因子的任何水平的平均值是否大于控制类别的平均值，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平均值是否相等。如果不是所有的平均值均相等，则检验一部分平均值的等同性。**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格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 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  $b$**  是排列组平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平均值。

**Scheffé** 检验的显著性水平可允许要检验的组平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，需要平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实测显著性水平。

**显示的检验。**为 LSD、Sidak、Bonferroni、Games-Howell、Tamhane T2 和 T3、Dunnett's C 以及 Dunnett T3 提供成对比较。为 S-N-K、Tukey  $b$ 、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's 的检验既是多重比较检验，同时也是多范围检验。

## GLM 估计边际均值

选择您需要的单元格中的总体边际平均值估计的因子和交互作用。为协变量（如果存在）调整这些平均值。

### 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平均值未修正的成对比较。只有在“显示以下项的平均值”列表选择了主效应的情况下，此项才可用。

### 比较简单主效应

只要目标列表包含一个或多个乘积或交互效应（例如， $A \times B$  和  $A \times B \times C$ ），就会启用该设置。该设置支持指定简单主效应之间的比较，其中，这些效应是嵌套在其他因子的级别内的主效应。

### 置信区间调整

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**比较主效应**和/或**比较简单主效应**的情况下才可用。

## 指定估计边际均值

1. 从 > 分析 > 一般线性模型下选择一种过程。
2. 在主对话框中，单击 **EM 平均值**。

## GLM 重复测量：保存

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。模型预测的因变量的值。
- 标准误差。对于自变量具有相同值的观测值所对应的因变量的均值的标准差的估计。

**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。

- 库克距离。这是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测量。较大的 Cook 距离表明从回归统计的计算中排除个案之后，系数会发生根本变化。
- 利用值。未居中的杠杆值。每个观测值对模型拟合度的相对影响。

**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。观测值和模型预测值之间的差值。
- 标准化。残差除以其标准差的估计值。标准化残差也称为 Pearson 残差，它的平均值为 0，标准差为 1。
- 学生化。残差除以其标准差的估计值，该估计值因个案而异，具体取决于自变量上每个个案的值与自变量均数的距离。有时称为内部学生化残差。
- 已删除。当某个观测值从回归系数的计算中排除时，该观测值的残差。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

此对话框中有一些可选统计。统计是使用固定效应模型计算的。

### 显示

#### 描述性统计

生成所有单元格中的所有因变量的观察平均值、标准差和计数。

#### 效应大小估算

给出每个效应和每个参数估计的部分  $\eta^2$  方值。 $\eta^2$  方统计描述总可变性中可归因于某个因子的部分。

#### 实测幂

基于实测值来设置备择假设时，获取该检验的功效。

#### 参数估算值

每个检验生成参数估计值、标准误差、 $t$  检验、置信区间和检验的观察势。

#### SSCP 矩阵

显示假设和误差 SSCP 矩阵。

#### 残差 SSCP 矩阵

显示残差 SSCP 矩阵。

#### 转换矩阵

显示残差协方差矩阵的巴特利特球形度检验。

### 同质性检验

生成跨主体间因子的所有水平组合的每个因变量的方差同质性的莱文检验（仅针对主体间因子）。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

用于检查有关数据的假设。如果不存在任何因子，那么禁用此项。

### 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。如果不存在任何因子，那么禁用此项。

### 失拟

检查因变量和自变量之间的关系能否由模型充分地描述。

### 一般可估函数

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都是一般可估计函数的线性组合。

### 显著性水平

您可能想要调整事后检验中使用的显著性水平和用于构建置信区间的置信水平。指定值还用于计算检验的检验效能。如果指定了显著性水平，那么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）。
- 为事后比较指定误差项（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）。
- 构造相关性矩阵数据文件（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）。

请参阅命令语法参考以获取完整的语法信息。

## 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，例如拆分图、单变量重复测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。

下面四种不同的方法可用于估计方差成分：最小范数二次无偏估计量 (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。

方差成分数据注意事项

**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平均值为零，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。

残差项的平均值也为零，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。来自不同实测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实测值是相关的。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

1. 从菜单中选择：

**分析 > 一般线性模型 > 方差成分...**

2. 选择一个自变量。

3. 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

**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。

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双阶交互。

**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。

**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。

## 构建项和定制项

### 构建项：

当您想要为一组选定因子和协变量的所有组合包含特定类型（例如主效应）的非嵌套项时，使用该选项。

### 构建定制项

如果要包含嵌套项，或者想要按变量显式构建任何项，请使用该选项。构建嵌套项包括以下步骤：

## 方差成分：选项

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量）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，那么此方法可生成所有无偏估计量的最小方差。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 (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。
- **受限最大似然 (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，因此其标准误差应比 ML 方法的标准误差要小。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测值具有相同的影响。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，那么方差估计值可用在“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。** 您可以指定收敛准则和最大迭代次数。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和类型。类型 III 最常用，而且是缺省类型。

**类型 I。** 此方法也称为平方和分层解构法。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双向交互效应之前指定，依此类推。
- 多项式回归模型，其中，任何低阶项都在任何高阶项之前指定。
- 纯嵌套模型，其中，指定的第一效应嵌套在指定的第二效应中，指定的第二效应嵌套在指定的第三效应中，依此类推。（此嵌套形式只能通过使用语法来指定。）

**类型 III。** 缺省值。此方法将设计中某个效应的平方和，计算为针对不含该效应的任何其他效应进行调整并与包含该效应的任何效应（如有）正交的平方和。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，平方和对于单元格频率就保持不变。因此，此类平方和常被认为对不带缺失单元格的不平衡模型有用。在不带缺失单元格的阶乘设计中，此方法等同于 Yates 均值加权平方方法。类型 III 平方和法常用于：

- “类型 I”中列出的所有模型。
- 没有空单元格的任何平衡或不平衡模型。

## 方差成分：保存到新文件

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-协方差矩阵或相关性矩阵保存到数据文件或数据集。仅当指定了**最大似然**或**受限最大似然**时才可用。

**创建值的文件。**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。主题。

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）。

请参阅命令语法参考以获取完整的语法信息。

## 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可变性。因此，混合线性模型可提供建模灵活性，不仅可以对数据的均值建模，还可以对其方差和协方差建模。

“线性混合模型”过程还是一种灵活的工具，适合于拟合可以规划为混合线性模型的其他模型。此类模型包括多级模型、分层线性模型和随机系数模型。

### 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) 和受限最大似然 (REML) 估计。

### 统计量

描述统计：各个不同的因子级别组合的因变量和协变量的样本大小、平均值和标准差。因子级别信息：每个因子级别及其频率的排序值。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。

## 线性混合模型数据注意事项

### 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该是定量的。主体和重复变量可以是任意类型。

### 假定情况

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平均值建模。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。重复度量对残差的协方差结构建模。此外，还假设该因变量来自正态分布。

### 相关过程

使用“探索”过程可以在运行分析之前检查数据。如果您并不认为存在相关或非常量可变性，您可使用“GLM 单变量”或“GLM 重复度量”过程。如果随机效应具有方差成分协方差结构，并且不存在重复测量，则可改用“方差成分分析”过程。

## 获取线性混合模型分析

1. 从菜单中选择：

### 分析 > 混合模型 > 线性...

2. (可选) 选择一个或多个主体变量。
  3. (可选) 选择一个或多个重复变量。如果定义重复变量, 请从下拉列表中选择**重复协方差类型**。
  4. (可选) 选择一个或多个 Kronecker 测量变量。
  5. (可选) 选择重复协方差类型。
  6. 单击**继续**。
  7. 选择一个自变量。
  8. 至少选择一个因子或协变量。
  9. 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: 主体与重复

此对话框可用于选择定义主体、重复观测值和 Kronecker 测量的变量, 以及选择残差的协方差结构。

### 主体

主体是可视为独立于其他主体的观察单元格。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。当每个主体都有重复测量, 而又要对这些观测值的相关性建立模型, 定义主体就变得尤为重要。例如, 您可预期同一个患者在连续多次就医时得到的血压读数是相关的。

主体也可以由多个变量的因子级别的组合来定义; 例如, 可以将性别和年龄类别指定为主体变量, 以将以下观点建模: 65 岁以上的男性彼此类似, 但与 65 岁以下的男性和女性不相关。

**主体**列表中指定的所有变量都会用于定义残差协方差结构的主体。您可以使用部分或全部变量来定义随机效应协方差结构的主体。

### 重复

此列表中指定的变量用于识别重复观测值。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)
- 直积 AR1(UN\_AR1)
- 直积非结构化 (UN\_UN)
- 复合对称直积 (UN\_CS)
- AR(1): 非齐性
- ARMA(1,1)
- 复合对称
- 复合对称: 相关性测量
- 复合对称: 异质
- 对角
- 因子分析: 一阶
- 因子分析: 一阶、异质
- Huynh-Feldt
- 标度恒等式
- Toeplitz
- Toeplitz: 非齐性
- 非结构化
- 未结构化: 相关性度量

- 空间：幂
- 空间：指数
- 空间：高斯
- 空间：线性
- 空间：线性对数
- 空间：球面

### 克罗内克测量

选择变量，用于指定 Knonecker 协方差测量的主体结构，以及确定如何关联测量误差。只有在选择下列其中一种**重复协方差类型**的情况下，此字段才可用。

- 直积 AR1(UN\_AR1)
- 直积非结构化 (UN\_UN)
- 复合对称直积 (UN\_CS)

### 空间协方差坐标

该列表中的变量指定为重复的协方差类型选择了其中一种空间协方差类型时重复观测值的坐标。

有关更多信息，请参阅第 77 页的『协方差结构』主题。

## 线性混合模型：固定效应

**固定效应。** 没有缺省模型，您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，那么可以排除截距。

**平方和。** 计算平方和的方法。对于不带缺失单元格的模型，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双阶交互。

**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

您可在此过程中为模型创建嵌套项。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。由于每位顾客只经常光顾某一位置的商店，因此可以说客户效应**嵌套在**商店位置效应中。

此外，还可以包含交互效应或将多层嵌套添加到嵌套项。

**限制。** 嵌套项存在下列限制：

- 交互效应中的所有因子必须是唯一的因子。因此，如果  $A$  是因子，则指定  $A*A$  是无效的。
- 嵌套效应中的所有因子必须是唯一的因子。因此，如果  $A$  是一个因子，那么指定  $A(A)$  是无效的。
- 任何效应都不能嵌套到协变量中。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，而且是缺省类型。



**类型 I。** 此方法也称为平方和分层解构法。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双向交互效应之前指定，依此类推。
- 多项式回归模型，其中，任何低阶项都在任何高阶项之前指定。
- 纯嵌套模型，其中，指定的第一效应嵌套在指定的第二效应中，指定的第二效应嵌套在指定的第三效应中，依此类推。（此嵌套形式只能通过使用语法来指定。）

**类型 III。** 缺省值。此方法将设计中某个效应的平方和，计算为针对不含该效应的任何其他效应进行调整并与包含该效应的任何效应（如有）正交的平方和。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，平方和对于单元格频率就保持不变。因此，通常认为此类平方和适用于不含缺失单元格的不平衡模型。在不含缺失单元格的阶乘设计中，此方法等同于 Yates 均值加权平方法。类型 III 平方和法常用于：

- “类型 I”中列出的所有模型。
- 没有空单元格的任何平衡或不平衡模型。

## 线性混合模型：随机效果

**协方差类型。** 此选项可用来指定随机效应模型的协方差结构。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)：非齐性
- ARMA(1,1)
- 复合对称
- 复合对称：相关性测量
- 复合对称：异质
- 对角
- 因子分析：一阶
- 因子分析：一阶、异质
- Huynh-Feldt
- 标度恒等式
- Toeplitz
- Toeplitz：非齐性
- 非结构化
- 未结构化：相关性度量
- 方差成分

有关更多信息，请参阅第 77 页的『协方差结构』主题。

**随机效应。** 没有缺省模型，您必须显式指定随机效应。还可以构建嵌套或非嵌套项。此外，还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。假设各个随机效应模型彼此无关；即，对于每个随机效应模型分别计算协方差矩阵。在同一随机效应模型中指定的项可以是相关的。

**主体组。** 列示的变量就是您在“选择主体/重复变量”对话框中选择用作主体变量的变量。请选择其中部分或全部变量，以便为随机效应模型定义主体。

**显示该组随机效应的参数预测。** 指定此项以显示随机效应参数估计。

## 线性混合模型：估算

### 方法

选择最大似然性或约束最大似然性。

### 自由度

提供用于为所有检验定义自由度的选项。

### 残差法

残差法将固定的自由度用于所有检验。当样本足够大、数据平衡或者模型使用较简单的协方差类型（例如，尺度化恒等或对角线）时，这种方法适用。

### Satterthwaite 近似法

Satterthwaite 法在检验之间使用场自由度。当样本很小、数据不平衡或模型使用复杂的协方差类型（例如，非结构化）时，这种方法适用。

### Kenward-Roger 近似法

Kenward-Roger 法针对固定效应参数的方差-协方差以及 t 检验和 F 检验中的近似分母自由度，提供更精确的小样本估计量。这种方法为 F 统计引入比例因子并对其进行估算，并使用 Taylor 级数展开来为数据中的估计随机结构引入分母自由度。

**注：**Kenward-Roger 法用于基于模型的协方差（而非鲁棒协方差）。在选择 Kenward-Roger 方法和稳健协方差时，Kenward-Roger 方法应用于基于模型的协方差，并且将显示以下警告：“因为选择 Kenward-Roger 方法，稳健协方差方法将更改为基于模型的协方差方法”。

### 迭代

可用选项如下：

#### 最大迭代数

请指定非负整数。

#### 最大逐步二分次数

每次迭代时，步长都会减去因子 0.5，直到对数似然估计增加或者达到最大逐步二分次数。指定正整数。

#### 每次达到以下步数打印一次迭代历史记录：n

显示一个表，在该表中，从第 0 次迭代（初始估计值）开始，每 n 次迭代就列出对数似然函数值和参数估计值。如果选择打印迭代历史记录，那么无论 n 值为多少，将总是打印最后一次迭代。

### 对数似然汇合。

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，那么假设收敛性。如果指定的值等于 0，那么不使用该准则。

### 参数汇合

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，则假定收敛性。如果指定的值等于 0，那么不使用该准则。

### Hessian 汇合。

对于**绝对**指定，如果基于 Hessian 的统计小于指定的值，则假定收敛性。对于**相对**指定，如果统计小于指定值与对数似然估计的绝对值的乘积，那么假设收敛性。如果指定的值等于 0，那么不使用该准则。

### 最大评分步数

请求将 Fisher 评分算法使用最多迭代数字 n。指定非负整数。

### 奇异性容差

这个值在检查奇异性时用作容差。指定一个正值。

## 线性混合模型：统计

汇总统计。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、平均值和标准差。这些统计量会针对因子的每个不同级别组合而显示。
- **个案处理摘要。** 显示因子、重复测量变量、重复测量主体以及随机效应主体及其频率的排序值。

模型统计。生成表：

- **固定效应的参数估计。** 显示固定效应参数估计及其大致标准误差。

- **协方差参数检验。** 显示协方差参数的渐近标准误差和 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性矩阵。
- **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。只有在至少指定一个随机效应的情况下，此选项才可用。如果为随机效应指定了主体变量，那么显示通用区组。
- **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函数。

**置信区间。** 每次构造置信区间时，都会使用此值。指定一个大于等于 0 且小于 100 的值。缺省值是 95。

## 线性混合模型：EM 均值

**拟合模型的估计边际均值。** 此组允许您请求单元格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差。此外，您还可以请求比较主效应的因子级别。

- **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。
- **显示下列各项的均值。** 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，那么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，那么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

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平均值相关的变量。

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

**预测值和残差。** 保存与模型拟合度值相关的变量。

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

## 线性混合模型-导出

此对话框允许您将 EBLUPs (经验最佳线性无偏预测) 输出表的内容导出到数据集或 .sav 文件。如果通过“随机”对话框指定了至少一个随机效应，并且选中了“显示此块的参数预测”框，那么将启用“线性混合模型”对话框中的“导出”按钮。

### 导出 EBLUPS

选择目标: **数据集** 或 **数据文件**

(可选) 提供名称。

如果指定了具有生成的 EBLUPs 的多个随机效应，请取消选中此复选框以将每个结果表显示在单独的数据集或文件中。

## 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）。

请参阅命令语法参考以获取完整的语法信息。

## 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。而且，该模型还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。** 航运公司可以使用广义线性模型拟合不同时间段建造的若干类型船舶的损坏计数的泊松回归，所得模型可帮助确定哪些船舶类型最容易损坏。

汽车保险公司可以使用广义线性模型来拟合汽车损坏索赔的伽马回归，所得模型可帮助确定对索赔金额贡献最大的因素。

医学研究人员可以使用广义线性模型来拟合区间删失生存数据的互补对数回归，以预测疾病复发时间。

广义线性模型数据注意事项

**数据。** 响应可以是标度、计数、二进制或试验中的事件。假设因子是分类型的。协变量、标度权重和偏移被假定为标度。

**假设。** 假设个案为独立实测值。

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...

1. 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）。
2. 在响应选项卡上，选择一个因变量。
3. 在预测变量选项卡上，选择用于预测因变量的因子和协变量。
4. 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链接函数，为按响应类型分类的几种常用模型提供快捷方式。

模型类型

**标度响应。** 可用选项如下：

- **线性。** 将正态指定为分布，将恒等式指定为关联函数。
- **具有对数链接的伽玛。** 将伽玛指定为分布，将对数指定为关联函数。

**有序响应。** 可用选项如下：

- **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。** 可用选项如下：

- **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。** 可用选项如下：

- **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。
- **区间删失生存**。将二项式指定为分布，将互补双对数指定为关联函数。

混合。可用选项如下：

- **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，将恒等式指定为关联函数。

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。存在许多可能的分布-关联函数组合，其中有许多组合适用于任何给定的数据集，因此，您可根据先验理论考虑事项或哪个组合看起来拟合得最好来指导您的选择。

- **二项式**。此分布仅适合表示二元响应或事件数量的变量。
- **伽玛**。该分布适用于具有正刻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刻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值可以为大于等于 0 的任何数字；您可以将其设置为固定值或者允许通过过程对其进行估算。辅助参数设置为 0 时，使用该分布等效于使用泊松分布。
- **正态**。该分布适合围绕某个中间值（平均值）呈对称钟型分布的标度变量。因变量必须是数值型变量。
- **泊松**。此分布可以被认为是在固定时间段内感兴趣的事件发生的次数，适用于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**。此分布适合由伽玛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**。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

关联函数是允许模型估计的因变量的转换。有下列函数可用：

- **Identity**.  $f(x)=x$ 。不会对因变量进行转换。此链接函数可用于任何分布。
- **Complementary log-log**.  $f(x)=\log(-\log(1-x))$ 。该函数只适用于二项分布。
- **累积概率**.  $f(x)=\tan(\pi(x-0.5))$ ，用于响应每个类别的累积概率。此函数仅适用于多项分布。
- **累积补充日志记录**.  $f(x)=\ln(-\ln(1-x))$ ，用于响应每个类别的累积概率。此函数仅适用于多项分布。
- **累积逻辑**.  $f(x)=\ln(x/(1-x))$ ，用于响应每个类别的累积概率。此函数仅适用于多项分布。
- **累积负面日志记录**.  $f(x)=-\ln(-\ln(x))$ ，用于响应每个类别的累积概率。此函数仅适用于多项分布。
- **累积 probit**.  $f(x)=\Phi^{-1}(x)$ ，应用于响应的每个类别的累积概率，其中  $\Phi^{-1}$  是逆标准正常累积分布函数。此函数仅适用于多项分布。
- **Log**.  $f(x)=\log(x)$ 。此链接函数可用于任何分布。
- **Log complement**.  $f(x)=\log(1-x)$ 。该函数只适用于二项分布。
- **Logit**.  $f(x)=\log(x/(1-x))$ 。该函数只适用于二项分布。
- **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。此函数仅适用于负二项式分布。
- **Negative log-log**.  $f(x)=-\log(-\log(x))$ 。该函数只适用于二项分布。
- **赔率**  $f(x)=[(x/(1-x))^{\alpha}-1]/\alpha$ ，如果  $\alpha \neq 0$ 。  $f(x)=\log(x)$ ，如果  $\alpha=0$ 。  $\alpha$  是必需的数字规范，而且必须是实数。该函数只适用于二项分布。

- **概率值。**  $f(x)=\Phi^{-1}(x)$ ，其中  $\Phi^{-1}$  是标准正态累积分布的反函数。该函数只适用于二项分布。
- **Power。**  $f(x)=x^\alpha$ ，if  $\alpha \neq 0$ 。  $f(x)=\log(x)$ ，如果  $\alpha=0$ 。 $\alpha$  是必需的数字规范，必须是实数。此链接函数可用于任何分布。

## 广义线性模型响应

在许多情况下，您可以只指定因变量；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，那么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。** 当响应是一组试验中发生的事件数时，因变量包含事件数，您可以选择包含试验数量的附加变量。或者，如果所有受试者的试验次数均相同，则可以使用固定值指定试验次数。试验次数应大于或等于每个案例的事件数。事件数量应为非负整数，试验次数应为正整数。

对于有序多项模型，您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。尺度权重是一些“已知”值，可能因观测值而异。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刻度参数除以该刻度权重变量。分析中不使用刻度权重值小于等于 0 或缺失的个案。

## 广义线性模型：参考类别

对于二元响应，可以为因变量选择参考类别。这会影响到某些输出，如参数估计值和保存的值，但不应更改模型拟合度。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，那么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，那么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

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量。

**因子。** 因子是分类预测变量，可以是数值或字符串。

**协变量。** 协变量为刻度预测变量，必须为数值。

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

**偏移量。** 偏移量项是“结构性”预测变量。它的系数不通过模型估计而假定其值为 1；因此，偏移量的值只与目标的线性预测变量简单相加。这对于泊松回归模型尤其有用，在这种模型中，每个观测值对于相关事件可以具有不同的揭示级别。

例如，对各个驾驶员的事故率建模时，三年驾驶经验的驾驶员在一次事故中的过错率与 25 年驾驶经验的驾驶员在一次事故中的过错率存在重大差别。如果将驾驶员的经验的自然对数作为偏移量项包括在内，那么可以按照泊松响应或负二项式响应事故数进行建模。

其他分布和关联类型的组合将需要偏移变量的其他转换。

## 广义线性模型：选项

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级别效应的值，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

### 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双阶交互。

#### 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

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

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

**限制：**嵌套项存在下列限制：

- 交互效应中的所有因子必须是唯一的因子。因此，如果  $A$  是一个因子，那么指定  $A*A$  是无效的。
- 嵌套效应中的所有因子必须是唯一的因子。因此，如果  $A$  是因子，则指定  $A(A)$  是无效的。
- 任何效应都不能嵌套到协变量中。因此，如果  $A$  是因子且  $X$  是协变量，则指定  $A(X)$  是无效的。

## 广义线性模型：估计

**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性，那么算法将继续执行 Newton-Raphson 方法。
- **标度参数方法。** 可以选择刻度参数估计方法。最大似然法可联合估计标度参数和模型效应；请注意，如果响应具有负二项式、泊松、二项式或多项分布，那么此选项无效。偏差和 Pearson 卡方选项从这些统计的值估计刻度参数。或者，可以为刻度参数指定固定值。
- **初始值。** 该过程将自动计算参数的初始值。另外，您也可以为参数估计指定初始值。

**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”的基于模型的估计量，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。** 可用选项如下：

- **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。

- **最大逐步二分法。** 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逐步二分次数。指定正整数。
- **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。

**收敛准则。** 下列选项可用：

- **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。** 对于“绝对值”指定，如果基于 Hessian 收敛性的统计小于指定的正值，那么假设收敛性。对于“相对”指定，如果统计小于指定的正值和对数似然估计的绝对值的乘积，那么假设收敛性。

**奇异值容差。** 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

如果指定了初始值，那么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，变量的顺序（从左到右）必须是：*RowType\_*、*VarName\_*、*P1*、*P2*、...，其中 *RowType\_* 和 *VarName\_* 是字符串变量，而 *P1*、*P2*、... 是对应于已排序参数列表的数字变量。

- 初始值在变量 *RowType\_* 的值为 *EST* 的记录中提供；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（如果响应具有多项分布）必须是列示的第一个初始值。
- 标度参数和负二项式参数（如果响应具有负二项分布）必须是指定的最后一个初始值。
- 如果“拆分文件”有效，那么变量必须以拆分文件变量开始（按创建“拆分文件”时指定的顺序），后跟 *RowType\_*、*VarName\_*、*P1*、*P2*、...，如上所述。拆分在指定数据集发生顺序必须与原始数据集中的顺序相同。主题。

注：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。最后一个参数后面的所有变量都将被忽略。最后一个参数之后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

**模型作用。** 可用选项如下：

- **分析类型。** 指定要生成的分析类型。类型 I 分析一般适用于您有先验理由来将模型中的预测变量排序的情况，类型 III 则适用于更一般的情况。Wald 或似然比统计量是根据卡方统计量组中的选择而计算的。
- **置信区间。** 指定一个大于 50 且小于 100 的置信水平。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。完整的函数包括一个相对于参数估计值来说不变的附加项；它对参数估计没有影响，在某些软件产品中不显示。

**打印。** 可用的输出如下所示：

- **观测值处理摘要。** 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、刻度权重变量、概率分布和关联函数。
- **拟合度统计量。** 显示离差和刻度化离差、Pearson 卡方和刻度化 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)。



- **模型汇总统计。** 显示模型拟合度检验，包括用于模型拟合度 Omnibus 检验的似然比统计，以及每种效应的类型 I 或 III 对比的统计。
- **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。
- **参数估计值的协方差矩阵。** 显示估计参数协方差矩阵。
- **参数估计值的相关性矩阵。** 显示估计参数相关性矩阵。
- **对比系数声明矩阵。** 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，并打印对梯度向量和海森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）开始每隔  $n$  次迭代就显示一次参数估计值，其中  $n$  代表打印区间的值。如果请求显示迭代历史记录，那么无论  $n$  为多少都会显示最后一次迭代。
- **标度参数或负二项式辅助参数的拉格朗日乘数检验。** 显示拉格朗日乘数检验统计量，用于评估使用偏差或皮尔逊卡方计算的标度参数的有效性，或设置为固定数字，用于正态分布、伽玛分布、逆高斯分布和特威迪分布。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

此选项卡用于显示因子级别和因子交互的估算边际平均值。您还可以要求显示整体估计平均值。估计边际平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。

**显示以下项的平均值。** 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级别。

- **成对。** 为指定或默示的因子的所有水平组合计算成对比较。这是因子交互效应唯一可用的对比方法。
- **简单。** 将每个水平的平均值与指定水平的平均值进行比较。当存在控制组时，此类对比很有用。
- **偏差。** 因子的每个水平与总平均值比较。偏移对比不是正交的。
- **差异。** 将每个水平的平均值（第一个水平除外）与前面水平的平均值进行比较。有时候将其称为逆 Helmert 对比。
- **Helmert。** 将因子的每个水平的均值（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。** 比较线性效应、二次效应、三次效应等等。第一自由度包含跨所有类别的线性效应；第二自由度包含二次效应，依此类推。这些对比常常用来估计多项式趋势。

**刻度。** 对于响应，可根据因变量的原始刻度计算估算边际平均值；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平均值。

**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較的显著性水平进行调节。使用此组可以选择调整方法。

- **最小显著性差异。** 此方法不控制拒绝某些线性对比与 null 假设值不同的假设的总体概率。
- **Bonferroni。** 此方法根据多项对比检验调整观察到的显著性水平。
- **连续 Bonferroni (Sequential Bonferroni)。** 这是一种连续下降的拒绝的 Bonferroni 过程，该过程在拒绝单个假设方面保守程度很低，但保持相同的整体显著性水平。
- **Sidak。** 此方法提供比 Bonferroni 方法更严格的界限。
- **连续 Sidak (Sequential Sidak)。** 这是一个逐步下降的排斥性 Sidak 过程，就排斥单个假设而言，其保守性小得多，且保持了相同的总体显著性水平。

## 广义线性模型：保存

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。

### 预测的响应平均值

以初始响应度规保存每个个案的模型预测值。当响应分布为二项式且因变量为二元时，该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。

### 响应平均值的置信区间的下限

保存响应平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。

### 响应平均值的置信区间上限

保存响应平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，这将保存每个案例的预测响应类别。此选项不支持其他响应分布。

### 线性预测变量的预测值

将每个案例的模型预测值保存在线性预测变量的指标中（通过指定的链接函数转换响应）。当响应分布是多项式时，该过程保存响应的每个类别（最后一个除外）的预测值，最多保存指定类别的数量。

### 线性预测变量的预测值的估计标准误差

当响应分布为多项式时，该过程会保存响应的每个类别（最后一个除外）的估计标准误差，最多可保存指定类别的数量。

如果响应分布是多项式，则以下各项不可用。

### 库克距离

这是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测量。较大的 Cook 距离表明从回归统计的计算中排除个案之后，系数会发生根本变化。

### 杠杆值

测量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  $(N-1)/N$ 。

### 残差

观测值和模型预测值之间的差值。

### Pearson 残差

某个案例对 Pearson 卡方统计的贡献的平方根（带原始残差的符号）。

### 标准化 Pearson 残差

对于此案例，则为 Pearson 残差乘以尺度参数和 1-杠杆的乘积的倒数的平方根。

### 偏差残差

某个案例对偏差统计的贡献的平方根（带原始残差的符号）。

### 标准化偏差残差

对于此案例，则为偏差残差乘以尺度参数和 1-杠杆的乘积的倒数的平方根。

### 似然残差

标准化 Pearson 残差和标准化偏差残差的平方的加权平均值（基于案例的杠杆）的平方根（带原始残差的符号）。

## 广义线性模型：导出

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、标准误差、显著性值和自由度的参数相关性或协方差矩阵。矩阵文件中变量顺序如下。

- **拆分变量**。定义拆分的任何变量（如果使用）。
- **RowType\_** 接受值（和值标签）**COV**（方差），**CORR**（相关性），**EST**（参数估计），**SE**（标准误差），**SIG**（显著性水平）和**DF**（采样设计自由度）。存在每个模型参数的 COV（或 CORR）行类型的单个个案，以及每个其他行类型的单个个案。
- **VarName\_** 接受对应于所有估算模型参数（刻度或负二项参数除外）的有序列表的值 P1 和 P2, ..., 对于行类型 COV 或 CORR，其值标签对应于参数估计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（必要时可包括刻度或负二项式参数）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；并且所有标准误差、显著性水平和残差自由度设为系统缺失值。

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

如果存在拆分，那么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、标准误差、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。使用该模型文件，可以将模型信息应用于其他数据文件以进行评分。

## 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）。
- 在计算估计边际平均值时修正值不同于其平均值的协变量（使用 EMMEANS 子命令）。
- 指定估计边际平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。

## 广义估算方程

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测量 Logistic 回归模型。

广义估计方程数据注意事项

**数据。** 响应可以是标度、计数、二进制或试验中的事件。假设因子是分类型的。协变量、标度权重和偏移被假定为标度。用于定义主体或主体内重复测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性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...

1. 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，但如果病人识别号在医院间不唯一，那么需要使用医院 ID 和病人 ID 的组合。在重复测量设置中，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。

2. 在**模型类型**选项卡上，指定分布和关联函数。
3. 在**响应**选项卡上，选择一个因变量。
4. 在**预测变量**选项卡上，选择用于预测因变量的因子和协变量。
5. 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，每个主体的重复测量因而按正确顺序在连续个案段中发生，那么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”的基于模型的估计量，即使错误地指定了工作相关性矩阵，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。** 此相关性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。** 重复测量不相关。
- **AR(1)。** 重复测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等于  $\rho$ ，对于被第三个元素分隔的元素等于  $\rho^2$ ，依此类推，约束为  $-1 < \rho < 1$ 。
- **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，由第三个测量分隔的测量对具有共同的相关系数，依此类推，直到由  $m-1$  个其他测量分隔的测量对。例如，您每年对从三年级到七年级的学生进行标准化考试。此结构假定第三年级和第四年级、第四年级和第五年级、第五年级和第六年级以及第六年级和第七年级的分数具有相同的相关性；第三年级和第五年级、第四年级和第六年级以及第五年级和第七年级具有相同的相关性；第三年级和第六年级以及第四年级和第七年级具有相同的相关性。假设间隔大于  $m$  的测量不相关。选择此结构时，请指定小于工作相关性矩阵的阶的  $m$  值。
- **未结构化。** 这是一个非常一般的相关性矩阵。

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，那么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。请指定非负整数。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。** 工作相关性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。如果工作相关性矩阵完全没有更新，则在整个估计过程中将使用初始工作相关性矩阵。如果该矩阵进行了更新，则可以指定更新工作相关性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。** 如果选中此项，那么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）的迭代之后停止。
- **Hessian 收敛性。** 如果基于 Hessian 的一个统计小于指定值（必须为正值），那么认为收敛性。

## 广义估计方程：模型类型

“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

标度响应。可用选项如下：

- **线性。** 将正态指定为分布，将恒等式指定为关联函数。
- **具有对数链接的伽玛。** 将伽玛指定为分布，将对数指定为关联函数。

有序响应。可用选项如下：

- **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。可用选项如下：

- **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。可用选项如下：

- **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。
- **区间删失生存**。将二项式指定为分布，将互补双对数指定为关联函数。

混合。可用选项如下：

- **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，将恒等式指定为关联函数。

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。存在许多可能的分布-关联函数组合，其中多个组合适用于任何给定的数据集，因此，您可根据先验理论考虑事项或哪个组合看起来拟合得最好来指导您的选择。

- **二项式**。此分布仅适合表示二元响应或事件数量的变量。
- **伽玛**。该分布适用于具有正刻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刻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值可以为大于等于 0 的任何数字；您可以将其设置为固定值或者允许通过过程对其进行估算。辅助参数设置为 0 时，使用该分布等效于使用泊松分布。
- **正态**。该分布适合围绕某个中间值（平均值）呈对称钟型分布的标度变量。因变量必须是数值型变量。
- **泊松**。此分布可以被认为是在固定时间段内感兴趣的事件发生的次数，适用于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**。此分布适合由伽玛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**。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。

链接函数

关联函数是允许模型估计的因变量的转换。有下列函数可用：

- **Identity**。  $f(x)=x$ 。不会对因变量进行转换。此链接函数可用于任何分布。
- **Complementary log-log**。  $f(x)=\log(-\log(1-x))$ 。该函数只适用于二项分布。
- **累积概率**。  $f(x)=\tan(\pi(x-0.5))$ ，用于响应每个类别的累积概率。此函数仅适用于多项分布。
- **累积补充日志记录**。  $f(x)=\ln(-\ln(1-x))$ ，用于响应每个类别的累积概率。此函数仅适用于多项分布。
- **累积逻辑**。  $f(x)=\ln(x/(1-x))$ ，用于响应每个类别的累积概率。此函数仅适用于多项分布。
- **累积负面日志记录**。  $f(x)=-\ln(-\ln(x))$ ，用于响应每个类别的累积概率。此函数仅适用于多项分布。
- **累积 probit**。  $f(x)=\Phi^{-1}(x)$ ，应用于响应的每个类别的累积概率，其中  $\Phi^{-1}$  是逆标准正常累积分布函数。此函数仅适用于多项分布。

- **Log.**  $f(x)=\log(x)$ 。此链接函数可用于任何分布。
- **Log complement.**  $f(x)=\log(1-x)$ 。该函数只适用于二项分布。
- **Logit.**  $f(x)=\log(x / (1-x))$ 。该函数只适用于二项分布。
- **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。此函数仅适用于负二项式分布。
- **Negative log-log.**  $f(x)=-\log(-\log(x))$ 。该函数只适用于二项分布。
- **赔率**  $f(x)=[(x/(1-x))^\alpha - 1]/\alpha$ ，如果  $\alpha \neq 0$ 。  $f(x)=\log(x)$ ，如果  $\alpha=0$ 。  $\alpha$  是必需的数字规范，而且必须是实数。该函数只适用于二项分布。
- **概率值.**  $f(x)=\Phi^{-1}(x)$ ，其中  $\Phi^{-1}$  是标准正态累积分布的反函数。该函数只适用于二项分布。
- **Power.**  $f(x)=x^\alpha$ ，if  $\alpha \neq 0$ 。  $f(x)=\log(x)$ ，如果  $\alpha=0$ 。  $\alpha$  是必需的数字规范，必须是实数。此链接函数可用于任何分布。

## 广义估计方程：响应

在许多情况下，您可以只指定因变量；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，那么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。** 当响应是一组试验中发生的事件数时，因变量包含事件数，您可以选择包含试验数量的附加变量。或者，如果所有受试者的试验次数均相同，则可以使用固定值指定试验次数。试验次数应大于或等于每个案例的事件数。事件数量应为非负整数，试验次数应为正整数。

对于有序多项模型，您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。尺度权重是一些“已知”值，可能因观测值而异。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刻度参数除以该刻度权重变量。分析中不使用刻度权重值小于等于 0 或缺失的个案。

## 广义估计方程：参考类别

对于二元响应，可以为因变量选择参考类别。这会影响到某些输出，如参数估计值和保存的值，但不应更改模型拟合度。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，那么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，那么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

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量。

**因子。** 因子是分类预测变量，可以是数值或字符串。

**协变量。** 协变量为刻度预测变量，必须为数值。

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

**偏移量。** 偏移量项是“结构性”预测变量。它的系数不通过模型估计而假定其值为 1；因此，偏移量的值只与目标的线性预测变量简单相加。这对于泊松回归模型尤其有用，在这种模型中，每个观测值对于相关事件可以具有不同的揭示级别。

例如，对各个驾驶员的事故率建模时，三年驾驶经验的驾驶员在一次事故中的过错率与 25 年驾驶经验的驾驶员在一次事故中的过错率存在重大差别。如果将驾驶员的经验的自然对数作为偏移量项包括在内，那么可以按照泊松响应或负二项式响应事故数进行建模。

其他分布和关联类型的组合将需要偏移变量的其他转换。

## 广义估计方程：选项

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级别效应的值，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

**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。或者，您也可以构建嵌套项或非嵌套项。

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双阶交互。

**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

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

此外，还可以包含交互效应或将多层嵌套添加到嵌套项。

**限制。** 嵌套项存在下列限制：

- 交互效应中的所有因子必须是唯一的因子。因此，如果  $A$  是因子，则指定  $A*A$  是无效的。
- 嵌套效应中的所有因子必须是唯一的因子。因此，如果  $A$  是因子，则指定  $A(A)$  是无效的。
- 任何效应都不能嵌套到协变量中。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

**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性，那么算法将继续执行 Newton-Raphson 方法。
- **刻度参数方法。** 可以选择刻度参数估计方法。

极大似然法可联合估计标度参数和模型效应；请注意，如果响应具有负二项式、泊松或二项分布，那么此选项无效。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；该标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标度参数。

偏差和 Pearson 卡方选项从初始广义线性模型中统计的值估计标度参数；然后，此标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，可为标度参数指定固定值。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。另外，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准则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，请参阅重复选项卡。

**迭代。** 可用选项如下：

- **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。
- **最大逐步二分法。** 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逐步二分次数。指定正整数。
- **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。

**收敛准则。** 下列选项可用：

- **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。** 对于“绝对值”指定，如果基于 Hessian 收敛性的统计小于指定的正值，那么假设收敛性。对于“相对”指定，如果统计小于指定的正值和对数似然估计的绝对值的乘积，那么假设收敛性。

**奇异值容差。** 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。处理相关性矩阵无需初始值，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。

如果指定了初始值，那么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，变量的顺序（从左到右）必须是：*RowType\_*、*VarName\_*、*P1*、*P2*、...，其中 *RowType\_* 和 *VarName\_* 是字符串变量，而 *P1*、*P2*、... 是对应于已排序参数列表的数字变量。

- 初始值在变量 *RowType\_* 的值为 *EST* 的记录中提供；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（如果响应具有多项分布）必须是列示的第一个初始值。
- 标度参数和负二项式参数（如果响应具有负二项分布）必须是指定的最后一个初始值。
- 如果“拆分文件”有效，那么变量必须以拆分文件变量开始（按创建“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。主题。

注：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。最后一个参数后面的所有变量都将被忽略。最后一个参数之后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

模型作用。可用选项如下：

- **分析类型。** 指定要模型效应检验生成的分析类型。类型 I 分析一般适用于您有先验理由来将模型中的预测变量排序的情况，类型 III 则适用于更一般的情况。根据在“卡方统计”组中的选择，计算 Wald 或一般化得分统计。
- **置信区间。** 指定一个大于 50 且小于 100 的置信水平。无论选择哪种卡方统计类型，Wald 区间总会生成，并且基于参数呈渐近正态分布的假设之上。
- **对数似然函数。** 它控制对数似然函数的显示格式。完整的函数包括一个相对于参数估计值来说不变的附加项；它对参数估计没有影响，在某些软件产品中不显示。

打印。显示下面的输出。

- **观测值处理摘要。** 显示分析以及“相关数据概要”表中所包括和排除的观测值的数量和百分比。
- **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。** 显示数据集名称、因变量或事件和试验变量、偏移变量、刻度权重变量、概率分布和关联函数。
- **拟合度统计量。** 显示 Akaike 信息标准的两个扩展以进行模型选择：用于选择最佳相关结构的 QIC 准则和另一个用于选择最佳预测变量子集的 QIC 度量。
- **模型汇总统计。** 显示模型拟合检验，其中包括模型拟合 Omnibus 检验的似然比统计量，以及每一种效应的类型 I 或 III 对比统计量。
- **参数估计。** 显示参数估计值和相应的检验统计和置信区间。除原始参数估计值外，还可以显示取幂的参数估计值。
- **参数估计值的协方差矩阵。** 显示估计参数协方差矩阵。
- **参数估计值的相关性矩阵。** 显示估计的参数相关性矩阵。
- **对比系数声明矩阵。** 为缺省效应和估计边际均值显示对比系数（如果“EM 平均值”选项卡上要求）。
- **常规可估计函数。** 显示生成对比系数声明矩阵的矩阵。
- **迭代历史记录。** 显示参数估计值和对数似然的迭代历史记录，并打印对梯度向量和海森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）开始每隔  $n$  次迭代就显示一次参数估计值，其中  $n$  代表打印区间的值。如果请求显示迭代历史记录，那么无论  $n$  为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

此选项卡用于显示因子级别和因子交互的估算边际平均值。您还可以要求显示整体估计平均值。估计边际平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。

**显示以下项的平均值。** 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级别。

- **成对。** 为指定或默示的因子的所有水平组合计算成对比较。这是因子交互效应唯一可用的对比方法。
- **简单。** 将每个水平的平均值与指定水平的平均值进行比较。当存在控制组时，此类对比很有用。
- **偏差。** 因子的每个水平与总平均值比较。偏移对比不是正交的。
- **差异。** 将每个水平的平均值（第一个水平除外）与前面水平的平均值进行比较。有时候将其称为逆 Helmert 对比。
- **Helmert。** 将因子的每个水平的均值（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。** 比较线性效应、二次效应、三次效应等等。第一自由度包含跨所有类别的线性效应；第二自由度包含二次效应，依此类推。这些对比常常用来估计多项式趋势。

**刻度。** 对于响应，可根据因变量的原始刻度计算估算边际平均值；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平均值。

**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较显著的显著性水平进行调节。使用此组可以选择调整方法。

- **最小显著性差异。** 此方法不控制拒绝某些线性对比与 null 假设值不同的假设的总体概率。
- **Bonferroni.** 此方法根据多项对比检验调整观察到的显著性水平。
- **连续 Bonferroni (Sequential Bonferroni).** 这是一种连续下降的拒绝的 Bonferroni 过程，该过程在拒绝单个假设方面保守程度很低，但保持相同的整体显著性水平。
- **Sidak.** 此方法提供比 Bonferroni 方法更严格的界限。
- **连续 Sidak (Sequential Sidak).** 这是一个逐步下降的排斥性 Sidak 过程，就排斥单个假设而言，其保守性小得多，且保持了相同的总体显著性水平。

## 广义估计方程：保存

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。

- **响应平均值的预测值。** 以初始响应度规保存每个个案的模型预测值。当响应分布为二项式且因变量为二元时，该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。
- **响应平均值的置信区间下限** 保存响应平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。
- **响应平均值的置信区间上限。** 保存响应平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，这将保存每个案例的预测响应类别。此选项不支持其他响应分布。
- **线性预测变量的预测值。** 将每个案例的模型预测值保存在线性预测变量的指标中（通过指定的链接函数转换响应）。当响应分布是多项式时，该过程保存响应的每个类别（最后一个除外）的预测值，最多保存指定类别的数量。
- **线性预测变量的预测值的估计标准误差。** 当响应分布为多项式时，该过程会保存响应的每个类别（最后一个除外）的估计标准误差，最多可保存指定类别的数量。

如果响应分布是多项式，则以下各项不可用。

- **原始残差。** 观测值和模型预测值之间的差值。
- **Pearson 残差。** 某个案例对 Pearson 卡方统计的贡献的平方根（带原始残差的符号）。

## 广义估计方程：导出

**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、标准误差、显著性值和自由度的参数相关性或协方差矩阵。矩阵文件中变量顺序如下。

- **拆分变量。** 定义拆分的任何变量（如果使用）。
- **RowType\_.** 接受值（和值标签）**COV**（方差），**CORR**（相关性），**EST**（参数估计），**SE**（标准误差），**SIG**（显著性水平）和**DF**（采样设计自由度）。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\_.** 接受对应于所有估算模型参数（刻度或负二项参数除外）的有序列表的值 **P1** 和 **P2, ...**，对于行类型 COV 或 CORR，其值标签对应于参数估计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（必要时可包括刻度或负二项式参数）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；并且所有标准误差、显著性水平和残差自由度设为系统缺失值。

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

如果存在拆分，那么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、标准误差、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。使用该模型文件，可以将模型信息应用于其他数据文件以进行评分。

## 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）。
- 指定固定工作相关性矩阵（使用 REPEATED 子命令）。
- 在计算估计边际平均值时修正值不同于其平均值的协变量（使用 EMMEANS 子命令）。
- 指定估计边际平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。

## 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各种模型，从简单线性回归模型到非正态纵向模型数据的复杂多级模型。

### 示例

地区教育委员会可以使用广义线性混合模型来判定一种实验教学方法对于提高数学成绩的有效性。由于来自同一班级的学生由同一个老师教导，因此这些学生是相关的，而且同一学校的各班级之间也是相关的，因此我们可以在学校和班级级别包含随机效应，以说明可变性的不同来源。

医学研究者可以用广义线性混合模型来判定一种新型抗惊厥药能否降低患者癫痫发作的比率。对同一患者重复测量通常是正相关的，因此适合使用具有随机效应的混合模型。目标字段为发作次数，其值为正整数，因此可能适合使用具有泊松分布和对数关联的广义线性混合模型。

电视、电话和网络服务的电缆供应商高管可以使用广义线性混合模型了解潜在的客户。由于可能的回答具有名义测量级别，因此公司分析人员使用具有随机截距的广义 logit 混合模型，以确定在给定调查响应者回答中不同服务类型（电视、电话、互联网）的服务使用情况回答之间的相关性。

“数据结构”选项卡用于在观测值具有相关性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测值，那么无需在此选项卡上指定任何设置。

## 效应选项

### 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，单个患者标识字段应足以在一个医院内定义主体，但如果患者标识号在医院间不唯一，那么需要使用医院标识和患者标识的组合。在重复测量设置中，将为每个主体记录多个观测值，因此每个主体可能在数据集中占用多条记录。

**主体**是可视作独立于其他主体的观测单元。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。当每个主体都有重复测量，而又要对这些观测值的相关性建立模型，定义主体就变得尤为重要。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。

在“变量”对话框上所有指定为**主体**的字段都用于定义残差协方差结构的主体，并提供可能字段的列表，用于定义随机效应块上的随机效应协方差结构主体。

### 重复测量

此处指定的字段用于标识重复观测值。例如，单个变量周可以标识医学研究中 10 周内的观测值，而月和天可共同用于标识一年内的每一天的观测值。

## 协方差选项

### 定义协方差组

此处指定的分类字段用于定义重复效应协方差参数的独立集合；每个由分组字段的交叉分类定义的类别都有一个集合。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。根据选择的**重复协方差类型**不同，提供的协方差选项也会有所不同。可用结构如下：

- 一阶自回归 (AR1)
- 直积 AR1 (UN\_AR1)
- 非结构化直积 (UN\_UN)
- 复合对称直积 (UN\_CS)
- 异质复合对称 (CSH)
- 异质自回归 (ARH1)
- 自回归移动平均值 (1,1) (ARMA11)
- 复合对称
- 对角
- 标度恒等
- Toeplitz
- 非结构化
- 方差成分
- 空间：幂
- 空间：指数
- 空间：高斯
- 空间：线性
- 空间：线性对数
- 空间：球形

### Kronecker 测量

选择变量，用于指定 Kronecker 协方差测量的主体结构，以及确定如何关联测量误差。只有在选择下列其中一种**重复协方差类型**的情况下，此字段才可用。

- 直积 AR1 (UN\_AR1)
- 非结构化直积 (UN\_UN)
- 复合对称直积 (UN\_CS)

### 空间协方差坐标

该列表中的变量指定为重复的协方差类型选择了其中一种空间协方差类型时重复观察的坐标。

有关更多信息，请参阅主题 [第 77 页的『协方差结构』](#)。

## 伪 R<sup>2</sup> 度量

伪 -R<sup>2</sup> 度量和类内相关系数包含在 GLMM 输出中 (当适用时)。伪 -R<sup>2</sup> 测量完全基于最终估计, 并在估计完成后生成。确定系数 R<sup>2</sup> 是一个通常报告的统计量, 因为它表示由线性模型解释的方差比例。类内相关系数 (ICC) 是一个相关统计, 用于量化由多级/分层数据中的分组 (随机) 因子解释的方差比例。

## 获取广义线性混合模型

此功能需要 SPSS Statistics Standard Edition 或“高级统计”选项。

从菜单中选择:

分析 > 混合模型 > 广义线性...

1. 在**数据结构**选项卡上定义数据集的主体结构。该选项卡提供多个用于指定模型效应和重复协方差类型的选项。
2. 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, 或者是事件/试验指定, 在此情况下, 事件和实验指定必须是连续的。还可以选择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。
3. 单击**构建选项**以指定可选的构建设置。
4. 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。
5. 单击**运行**以运行过程并创建模型对象。

## 目标

这些设置用于定义目标及其分布, 以及通过关联函数来定义目标与预测变量的关系。

**目标。** 目标是必需的。它可以具有任何测量级别, 目标的测量级别将限制哪些分布和关联函数是适当的。

- **使用试验次数作为分母。** 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, 在试验一种新型杀虫剂时, 您可以将蚂蚁样本投入不同浓度的杀虫剂中, 然后记录被杀死的蚂蚁数量和每个样本的蚂蚁数量。在本例中, 用于记录被杀死的蚂蚁数量的字段应该指定为目标 (事件) 字段, 用于记录每个样本中的蚂蚁数量的字段应该指定为试验字段。如果每个样本的蚂蚁数量相同, 那么可以使用固定值来指定试验次数。

对于每条记录, 试验次数应大于或等于事件数量。事件数量应为非负整数, 试验次数应为正整数。

- **自定义参考类别。** 对于分类目标, 您可以选择参考类别。这可能会影响某些输出 (例如, 参数估计值), 但是应该不会更改模型拟合。例如, 如果目标的值为 0、1 和 2, 那么缺省情况下, 该过程会将最后一个类别 (值最大的类别, 也就是本例中的 2) 作为参考类别。在这种情况下, 参数估计应解释为与类别 0 或 1 的似然估计相关, 而该似然估计相对于类别 2 的似然估计。如果您指定定制类别, 并且您的目标已经定义了标签, 那么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, 当您不准确记得特定字段的编码方式时, 这会很方便。

**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, 模型将期望目标的值的分布遵循所指定的形状, 并且目标值通过所指定的关联函数与预测变量线性相关。针对一些常用模型提供了快捷方式, 如果您要拟合的特定分布与关联函数组合不在此快捷列表上, 也可以选择**定制**设置。

- **线性模型。** 指定带有恒等关联的正态分布, 在可以使用线性回归或者 ANOVA 模型来预测目标时很有用。
- **伽玛回归。** 指定带有对数关联的伽马分布, 当目标包含所有正值, 且偏向较大值时, 应使用该分布。
- **对数线性。** 指定带有对数关联的泊松分布, 当目标表示固定时间段内出现的次数时应该使用该分布。
- **负二项回归。** 使用对数关联函数指定负二项式分布, 在目标和分母代表观察到第  $k$  次成功所需的试验次数时使用。
- **多项 Logistic 回归。** 指定多项分布, 当目标是多类别响应时, 应该使用多项分布。它使用累积 Logit 关联 (有序结果) 或者广义 Logit 关联 (多类别名义响应)。
- **二元 logistic 回归。** 指定带有 logit 关联的二项分布, 当目标是逻辑回归模型所预测的二元响应时, 应该使用该分布。
- **二元 Probit。** 指定带有概率值关联的二项分布, 当目标是具有基础正态分布的二元响应时, 应该使用该分布。

- **区间删失生存。** 指定带有互补对数-对数关联的二项分布，当一些观察没有终止事件时，这在生存分析中很有用。

## 分布

此选择指定目标的分布。能够指定非正态分布和非恒等相关函数是广义线性混合模型相对于线性混合模型的重大改进。存在许多可能的分布-关联函数组合，其中有若干组合可适用于任何给定的数据集，因此您可以根据先验理论考虑的事项或哪个组合看起来拟合得最好来指导您的选择。

- **二项式。** 此分布仅适合于表示二元响应或事件数量的目标。
- **伽玛。** 此分布适合于具有偏向较大正值的正比例值的目标。如果数据值小于或等于 0 或者缺失，那么在分析中不会使用相应的个案。
- **逆高斯。** 此分布适合于具有偏向较大正值的正比例值的目标。如果数据值小于或等于 0 或者缺失，那么在分析中不会使用相应的个案。
- **多项式。** 该分布适合于表示多类别响应的目标。模型的形式将取决于目标的测量级别。

名义目标将产生名义多项模型，其中针对目标的每个类别（参考类别除外）单独估计一组模型参数。所给定预测变量的参数估计值显示了该预测变量与目标的每个类别相对于参考类别的可能性之间的关系。

一个有序目标会返回一个有序多项模型，其中的传统截距项会由一组与目标类别累积概率相关的**阈值参数**集替换。

- **负二项式。** 负二项回归使用带有对数关联的负二项分布，当目标表示具有较大方差的出现次数时，应该使用该分布。
- **正态。** 这适合于其值围绕中心值（平均值）呈对称的钟形分布的连续目标。
- **泊松。** 此分布可以被认为是在固定时间段内感兴趣的事件发生的次数，适用于具有非负整数值的变量。如果数据值不是整数、小于 0 或者缺失，那么在分析中不会使用相应的个案。

## 关联函数

关联函数是对目标的一种变换，允许对模型进行估计。有下列函数可用：

- **Identity.**  $f(x)=x$ 。不对目标进行变换。此关联可用于除了多项式分布之外的任何分布。
- **Complementary log-log.**  $f(x)=\log(-\log(1-x))$ 。这仅适合于二项式分布或多项式分布。
- **Cauchit.**  $f(x) = \tan(\pi(x - 0.5))$ 。这仅适合于二项式分布或多项式分布。
- **Log.**  $f(x)=\log(x)$ 。此关联可用于除了多项式分布之外的任何分布。
- **Log complement.**  $f(x)=\log(1-x)$ 。这仅适合于二项式分布。
- **Logit.**  $f(x)=\log(x / (1-x))$ 。这仅适合于二项式分布或多项式分布。
- **Negative log-log.**  $f(x)=-\log(-\log(x))$ 。这仅适合于二项式分布或多项式分布。
- **概率值.**  $f(x)=\Phi^{-1}(x)$ ，其中  $\Phi^{-1}$  是标准正态累积分布的反函数。这仅适合于二项式分布或多项式分布。
- **Power.**  $f(x)=x^\alpha$ , if  $\alpha \neq 0$ .  $f(x)=\log(x)$ ，如果  $\alpha=0$ 。 $\alpha$  是必需的数字规范，必须是实数。此关联可用于除了多项式分布之外的任何分布。

## 固定效应




固定效应因子通常被认为是这样的字段：其感兴趣的值全部都在数据集中表示，并且可以用于评分。缺省情况下，会在模型的固定效应部分输入具有未在对话框中其他位置指定的预定义输入角色的字段。分类（名义和有序）字段用作模型中的因子，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中，来将效应输入到模型。所创建效应的类型取决于您将选择放下的热点。

- **主。** 放下的字段显示为效应列表底部单独的主效应。
- **双向。** 所有可能的放下字段配对显示为效应列表底部的双向交互。
- **三阶。** 放下字段的所有可能三元组显示为效应列表底部的三向交互。
- **\*** 所有放下字段的组合显示为效应列表底部的单个交互。

通过“效应构建器”右侧的按钮可以执行各种操作。

表 1: 效应构建器按钮描述

图标	描述
	通过选择要删除的项，然后单击“删除”按钮来删除固定效应模型中的项。
	通过选择要重新排序的项，然后单击向上箭头或向下箭头来对固定效应模型中的项进行重新排序。
	使用第 41 页的『添加自定义项』对话框并通过单击“添加定制项”按钮，可以向模型添加嵌套项。

**包括截距。** 截距通常会包括在模型中。如果您可以假设数据穿过原点，那么可以排除截距。

## 添加自定义项

您可以在此过程中为模型构建嵌套项。对于对其值与另一个因子的级别不交互的因子或协变量的效应进行建模，嵌套项很有用。例如，一家杂货连锁店可能在不同商店位置迎合顾客的不同消费习惯。由于每位顾客只经常光顾某一位置的商店，因此可以说客户效应嵌套在商店位置效应中。

此外，您可以包括交互效应（例如，涉及相同协变量的多项式项），或者向嵌套项添加多级嵌套。

**限制。** 嵌套项存在下列限制：

- 交互中的所有因子必须唯一。因此，如果  $A$  是一个因子，那么指定  $A*A$  是无效的。
- 嵌套效应中的所有因子必须唯一。因此，如果  $A$  是因子，则指定  $A(A)$  是无效的。
- 任何效应都不能嵌套在协变量中。因此，如果  $A$  是因子且  $X$  是协变量，则指定  $A(X)$  是无效的。

构造嵌套项

1. 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。
2. 单击（内部）。
3. 选择嵌套了前一个因子或协变量的因子，然后单击箭头按钮。
4. 单击添加项。

（可选）您可以包括交互效应，或者向嵌套项添加多级嵌套。

## 随机效应

随机效应因素是一些字段，它们在数据文件中的值可视为来自较大值群的随机样本。它们对于解释目标的过度可变性很有用。缺省情况下，如果您在“数据结构”选项卡中选择了多个主题，那么会对最内层主题之外的每个主题创建一个随机效应块。例如，如果您选择了学校、班级和学生作为“数据结构”选项卡上的主题，那么会自动创建下列随机效应块：

- 随机效应 1：主题是学校（没有效应，只影响截距）
- 随机效应 2：主题是学校 \* 班级（没有效应，只影响截距）

您可以通过下列方式使用随机效应块：

1. 要添加新的块，请单击**添加块...**这将打开第 41 页的『随机效应块』对话框。
2. 要编辑现有块，请选择要编辑的块，然后单击**编辑块...**这将打开第 41 页的『随机效应块』对话框。
3. 要删除一个或多个块，请选择要删除的块，然后单击“删除”按钮。

## 随机效应块

通过在源列表中选择一个或多个字段并将其添加至**效应构建器**列表，可将效应输入到模型中。

创建的效应类型取决于您所选择的**类型**。分类（名义和有序）字段用作模型中的因子，连续字段用作协变量。

#### **交互**

所有字段的组合显示为效应列表底部的单个交互。

#### **主效应**

放下的字段显示为效应列表底部单独的主效应。

#### **所有二阶**

所有可能的放下字段配对显示为效应列表底部的双向交互。

#### **所有三阶**

放下字段的所有可能三元组显示为效应列表底部的三向交互。

#### **所有四阶**

放下字段的所有可能三元组显示为效应列表底部的四向交互。

#### **所有五阶**

放下字段的所有可能三元组显示为效应列表底部的五向交互。

#### **包括截距**

缺省情况下，随机效应模型不包含截距。如果您可以假设数据穿过原点，那么可以排除截距。

#### **显示此块的参数预测**

指定此项以显示随机效应参数估计。

#### **主体组合**

这允许您从“变量”对话框的预设置主题组合中指定随机效应主题。例如，在“变量”对话框上，如果依次将学校、班级和学生定义为主体，那么“主体组合”下拉列表将会包含**无**、**学校**、**学校 \* 班级**和**学校 \* 班级 \* 学生**选项。

#### **随机效应协方差类型**

这指定残差的协方差结构。可用的结构为：

- 一阶自回归 (AR1)
- 自回归移动平均 (1,1) (ARMA11)
- 复合对称
- 对角线
- 比例
- 托普利兹
- 非结构化
- 方差分量

#### **定义协方差组的依据**

此处指定的分类字段将定义独立的随机效应协方差参数集；由分组字段的交叉分类所定义的每个类别各有一个分类字段。可以为每个随机效应块指定一组不同的分组字段。所有受试者具有相同协方差类型；相同协方差分组内的受试者具有相同的参数值。

## **权重和偏移量**

**分析权重。** 刻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，可能会因观察而异。如果指定了“分析权重”字段，那么刻度参数（与响应方差相关）都会除以每个观测的分析权重值。分析中不会使用分析权重值小于等于 0 或缺失的记录。

**偏移量。** 偏移量项是“结构性”预测变量。它的系数不通过模型估计而假定其值为 1；因此，偏移量的值只与目标的线性预测变量简单相加。这对于泊松回归模型尤其有用，在这种模型中，每个观测值对于相关事件可以具有不同的揭示级别。

例如，对各个驾驶员的事故率建模时，三年驾驶经验的驾驶员在一次事故中的过错率与 25 年驾驶经验的驾驶员在一次事故中的过错率存在重大差别。如果将驾驶员的经验的自然对数作为偏移量项包括在内，那么可以按照泊松响应或负二项式响应事故数进行建模。

其他分布和关联类型的组合将需要偏移变量的其他转换。



## 常规构建选项

这些选择指定一些用于构建模型的更高级条件。

### 排列顺序

这些控件用于确定目标和因子（分类输入）类别的顺序，以确定“最后一个”类别。如果目标未分类或者在第 39 页的『目标』设置上指定了定制参考类别，那么目标排序顺序设置将被忽略。

### 停止规则

您可指定算法执行的最大迭代次数。此算法使用包含内层循环和外层循环的双迭代式过程。指定的最大迭代次数值适用于这两个循环。请指定非负整数。缺省值为 100。

### 估计后设置

这些设置确定如何计算一些用于查看的模型输出。

#### 置信度级别 (%)

这是用于计算模型系数的区间估计值的置信度级别。请指定大于 0 且小于 100 的值。缺省值为 95。

#### 自由度

用来指定如何计算显著性检验的自由度。如果样本够大、数据平衡或模型使用较简单的协方差类型（例如，调整恒等式或对角线），可选择**残差方法**。这是缺省设置。如果样本很小、数据不平衡或模型使用复杂的协方差类型（例如非结构化），请选择 **Satterthwaite 近似法**。如果样本较小，并且您使用的是受限最大似然 (REML) 模型，请选择 **Kenward-Roger 近似**。

#### 固定效应和系数检验

这是用于计算参数估计值协方差矩阵的方法。如果担心实验数据违反模型假设，可选择稳健估计。

## 估计

模型构建算法使用包含内层循环和外层循环的双重迭代式过程。下列设置适用于内层循环。

### 参数收敛。

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，则假定收敛性。如果指定的值等于 0，那么不使用该准则。

### 对数似然收敛。

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，那么假设收敛性。如果指定的值等于 0，那么不使用该准则。

### Hessian 收敛。

对于**绝对**指定，如果基于 Hessian 的统计小于指定的值，则假定收敛性。对于**相对**指定，如果统计小于指定值与对数似然估计的绝对值的乘积，那么假设收敛性。如果指定的值等于 0，那么不使用该准则。

### 最大 Fisher 评分步骤数。

请指定非负整数。值为 0 则指定 Newton-Raphson 方法。如果值大于 0，那么指定最多使用迭代次数  $n$  次 Fisher 评分算法，其中  $n$  是指定的整数，其后是 Newton-Raphson。

### 奇异性容差。

这个值在检查奇异性时用作容差。指定一个正值。

注: 缺省情况下使用“参数收敛”，此时，将选中容差为 1E-6 的最大**绝对**变化。此设置可能会生成与 V22 之前的版本中获得的结果不同的结果。要重新生成版本 22 之前的版本中的结果，请对参数收敛准则使用**相对**，并保留缺省容差值 1E-6。

## 估计均值

此选项卡用于显示因子级别和因子交互的估算边际平均值。多项模型没有估算边际均值。

项。固定效应的模型项完全由此处列出的类别字段组成。检查需要模型生成估计边际均值的每一个项。

- **对比类型**。指定用于对比字段级别的对比类型。如果选择**无**，将不生成对比。**成对**可生成指定因子所有级别组合的成对比较。这是因子交互效应唯一可用的对比方法。**偏差**对比将因子的每个级别与总平均值进行比较。**简单**对比是因子的每个级别（最后一个除外）与最后一个级别的比较。“最后一个”级别由“构建选项”上指定的因子排序顺序决定。请注意，此处所有对比类型均为非正交。

- **对比字段。** 指定一个因子，该因子的级别将使用所选对比类型进行比较。如果选择无作为对比类型，那么不能（或无需）选择任何对比字段。

**连续字段。** 所列连续字段是从使用连续字段的固定效应项中抽出的。计算估算边际平均值时，协方差固定为指定的值。请选择平均值或指定定制值。

**显示估计平均值依据。** 用于指定是否对基于目标原始尺度或关联函数变换的估算边际均值进行计算。**原始目标刻度**计算目标的估计边际平均值。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）的估计边际平均值。**关联函数转换**计算线性预测变量的估计边际均值。

**多重比较的调整方法。** 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较显著的显著性水平进行调节。这允许您选择调节方法。

- **最小显著性差异。** 此方法不控制拒绝某些线性对比与 null 假设值不同的假设的总体概率。
- 连续 *Bonferroni (Sequential Bonferroni)*. 这是一种连续下降的拒绝的 Bonferroni 过程，该过程在拒绝单个假设方面保守程度很低，但保持相同的整体显著性水平。
- 连续 *Sidak (Sequential Sidak)*. 这是一个逐步下降的排斥性 Sidak 过程，就排斥单个假设而言，其保守性小得多，且保持了相同的总体显著性水平。

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

## 保存

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。缺省字段名称是 *PredictedValue*。

**分类目标的预测概率-** 如果目标是分类目标，那么此关键字将保存前  $n$  个类别的预测概率，直至指定为 **要保存的最大类别数** 的值。计算出的值为有序目标的累积概率。缺省根名称是 *PredictedProbability*。要保存预测类别的预测概率，可以保存置信度（参阅下文）。

**置信区间-** 为预测值或预测概率保存置信区间的上限和下限。对于除多项分布之外的所有分布，这将创建两个变量，缺省根名称是 *CI*，后缀分别为 *\_Lower* 和 *\_Upper*。

对于多项分布和名义目标，为每个因变量类别创建一个字段。这将保存前  $n$  个类别的预测概率的下限和上限，直到指定为 **要保存的最大类别数** 的值。缺省根名称是 *CI*，缺省字段名称是 *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。

对于多项分布和有序目标，为每个因变量类别创建一个字段，但最后一个类别除外（有关更多信息，请参阅第 43 页的『常规构建选项』主题）。这将保存前  $n$  个类别（直到但不包括最后一个类别）的累积预测概率的下限和上限，直到指定为 **要保存的最大类别数** 的值。缺省根名称是 *CI*，缺省字段名称是 *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-** 为每条记录保存 Pearson 残差，可用于模型拟合的估算后诊断。缺省字段名称是 *PearsonResidual*。

**置信度-** 保存分类目标的预测值中的置信度。计算的置信度可基于预测值的概率（最高的预测概率）或最高预测概率和次高预测概率之间的差异。缺省字段名称是 *Confidence*。

## 导出

**导出模型。** 这会将模型写入到外部 *.zip* 文件。使用该模型文件，可以将模型信息应用于其他数据文件以进行评分。请参阅主题 [评分向导](#) 以获取更多信息。请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。

### 导出输出

**导出 EBLUPS。** 如果通过“字段和效应”选项卡上的“随机效应”项对话框创建了至少一个随机效应块，并且选中了用于显示该块的参数预测的复选框，那么将启用此部分。如果指定了多个此类块，那么可以选择将结果保存到单独的数据集或文件中，或者将结果组合到单个数据集或文件中。

## 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。

缺省情况下，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。

作为模型对象的替代方法，可以通过在“选项”对话框（编辑 > 选项）的“输出”选项卡上的“输出显示”组中选择**透视表和图表**来生成透视表和图表。以下主题进一步提供模型对象的描述。

### 模型摘要

此视图是模型及其拟合的快照和概览摘要。

**表。**表可标识目标、概率分布和在目标设置上指定的关联函数。如果目标是由事件和试验定义，那么单元格将会拆分，以显示事件字段以及试验字段或者固定次数的试验。此外，还会显示经有限样本校正的 Akaike 信息准则 (AICC) 和 Bayesian 信息准则 (BIC)。

- *Akaike* 已校正。用于根据  $-2$  (受限) 对数似然选择和比较混合模型的度量。值越小，模型越好。AICC“修正”适用于小样本的 AIC。随着样本大小增加，AICC 收敛性为 AIC。
- *Bayesian*。用于根据  $-2$  对数似然选择和比较模型的度量。值越小，模型越好。BIC 也会“惩罚”过多参数模型（例如，具有大量输入的复杂模型），但比 AIC 更严格。

**图。**如果目标为分类目标，那么图表将显示最终模型的准确性，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测量。将为每个主体字段和重复测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测量字段的级别数。

### 预测-实测

对于连续目标（包括指定为事件/试验的目标），这将显示已分箱的散点图，其中，在纵轴上显示预测值，在横轴上显示实测值。理想情况下，这些点应该位于 45 度线上；此视图可以告诉您是否有任何记录被模型预测得特别糟糕。

### 分类

对于分类目标，这将显示热图中观测值与预测值的交叉分类以及整体正确百分比。

**表样式。**有多种不同的显示样式，可以从**样式**下拉列表中进行访问。

- **行百分比。**此样式将在单元格中显示行百分比（单元格计数表示为行总数的百分比）。这是缺省选项。
- **单元格计数。**这将显示单元格中的单元格计数。热图的阴影部分仍然基于行百分比。
- **热图。**此样式将只显示阴影，而不会在单元格中显示值。
- **压缩。**此样式不会在单元格中显示行或列标题，也不会显示值。在目标具有许多分类时，此样式将十分有用。

**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有任何影响。

**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。**如果所显示的目标有 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**下拉列表中进行访问。

- **图。** 此图表中的效应按照“固定效应”设置中指定的顺序从上到下排序。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）。这是缺省选项。
- **表。** 这是 ANOVA 表，其中包含总体模型与各个模型效应。效应按照其在固定效应设置里指定的顺序从上到下排序。

**显著性。** 这里有一个“显著性”滑块，用于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。这不会更改模型，而只是让您关注最重要的系数。缺省情况下此值为 1.00，因此不会根据显著性来过滤系数。

## 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**下拉列表中进行访问。

- **图。** 此图表首先显示截距，然后按照“固定效应”设置中指定的顺序从上到下对效应进行排序。在包含因子的效应中，系数按数据值升序进行排序。在图表中，连接线的颜色和粗细取决于系数的显著性，粗线对应于较显著的系数（ $p$  值较小）。这是缺省样式。
- **表。** 这将显示各个模型系数的值、显著性检验和置信区间。在截距之后，效应按照其在固定效应设置里指定的顺序从上到下排序。在包含因子的效应中，系数按数据值升序进行排序。

**多项式。** 如果多项分布起作用，那么多项式下拉列表将控制要显示的目标分类。这些值在列表中的排序顺序由“构建选项”设置上的规范决定。

**指数。** 用于显示某些模型类型的指数分布系数估计值和置信区间，包括二元 Logistic 回归（二项式分布和分对数关联函数），名义 Logistic 回归（多项式分布和分对数关联函数），负二项式回归（负二项式分布和分对数关联函数）以及对数线性模型（泊松分布和分对数关联函数）。

**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隐藏。这不会更改模型，而只是让您关注最重要的系数。缺省情况下此值为 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**下拉列表中进行访问。

- **协方差值。** 这是协方差矩阵热图，效应在其中按照“固定效应”设置中指定的顺序从上到下进行排序。corrgram 图中的颜色对应于键中所显示的单元格值。这是缺省选项。
- **相关图。** 这是协方差矩阵的热图。
- **压缩。** 这是没有行和列标题的协方差矩阵热图。

**区组。** 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。** 如果随机效应块具有组规范，那么会有“组”下拉列表，用于选择要显示的组级别。

**多项式。** 如果多项分布起作用，那么多项式下拉列表将控制要显示的目标分类。这些值在列表中的排序顺序由“构建选项”设置上的规范决定。

## 协方差参数

此视图显示协方差参数估计值，还有残差和随机效应的相关统计。这些是高级而重要的结果，提供协方差结构是否合适的信息。

**摘要表。** 这是一个关于残差（**R**）和随机效应（**G**）协方差矩阵的参数数量，固定效应（**X**）和随机效应（**Z**）设计矩阵中的排序（列数）以及由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值、标准误差和置信区间。所显示的参数数量取决于效应的协方差结构，对于随机效应块，那么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区组是分组指定的，那么会有“组”下拉列表来选择要显示的分组级别。

**多项式。** 如果多项分布起作用，那么多项式下拉列表将控制要显示的目标分类。这些值在列表中的排序顺序由“构建选项”设置上的规范决定。

## 估计均值：显著效应

这些是为 10 个“最显著”的固定全因子效应显示的图表，首先显示的是三向交互，然后显示的是双向交互，最后显示的是主效应。该图表在横轴上显示主效应（或者交互中第一个列出的效应）的每个值的纵轴上目标的模型估计值；为交互中第二个列出的效应的每个值生成单独的折线；对于三向交互中第三个列出的效应的每个值，将生成单独的图表；所有其他预测变量保持不变。它提供了目标上每个预测变量的系数的效应的有用可视化显示。请注意，如果没有显著的预测变量，那么不会生成估计均值。

**置信度。** 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。

## 估计平均值：定制效应

这些是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**下拉列表中进行访问。

- **图。** 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级别；对于交互，将为对比字段之外的每个效应级别组合显示图表。对于**成对对比**，会是一个距离网络图表；也就是图形化表示的比较表，表中网络节点间的距离与样本间的差异相对应。黄线对应于统计上的显著差异；黑线对应于不显著的差异。将鼠标悬停在该网络中的某条线上，可显示该线所连节点间的调整差异显著性工具提示。

对于**偏差对比**，会显示目标模型估计值位于垂直轴上，而对比字段值位于水平轴上的条形图；对于交互效应，每个效应级别组合（而非对比字段）会显示一个图表。这些条形图将以黑色水平线表示对比字段的每个水平和总平均值之间的差异。

对于**简单对比**，会显示目标模型估计值位于垂直轴上，而对比字段值位于水平轴上的条形图；对于交互效应，每个效应级别组合（而非对比字段）会显示一个图表。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、其标准误差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、标准误差、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。

**置信度。** 这将切换使用指定为“构建选项”一部分的置信度的边际平均值的置信上限和下限的显示。

**布局。** 用于切换对对比图表的布局。圆形布局不如网状布局能使对比明显，但可避免重叠线。

## 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度到多维交叉制表。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格加上 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。** 频率、残差、参数估计值、标准误差、置信区间和偏关联检验。对于定制模型，那么为残差图和正态概率图。

## 模型选择对数线性分析数据注意事项

**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字变量。

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格具有少量的观测值，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。如果数字变量具有空类别，那么使用“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...

1. 选择两个或更多数值型分类因子。
2. 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**。
3. 定义每个因子变量的值范围。
4. 在“建立模型”组中选择一个选项。

或者，您可以选择一个单元格权重变量指定结构零。

## 对数线性分析：定义范围

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，那么只使用 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

**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 3 路交互  $A*B*C$ ，双向交互  $A*B$ 、 $A*C$  和  $B*C$  以及 A、B 和 C 的主要影响。请勿在生成类中指定较低顺序的亲属。

对于选定因子：

### 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双阶交互。

### 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。

## 构建项和定制项

### 构建项：

当您想要为一组选定因子和协变量的所有组合包含特定类型（例如主效应）的非嵌套项时，使用该选项。

## 构建定制项

如果要包含嵌套项，或者想要按变量显式构建任何项，请使用该选项。构建嵌套项包括以下步骤：

## 模型选择对数线性分析：选项

**显示。** 您可以选择**频率**和/或**残差**。在饱和模型中，实测和期望的频率相同，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**、**收敛性**或 **Delta**（为饱和模型的所有单元格频率添加的值）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格权重（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）。

请参阅命令语法参考以获取完整的语法信息。

## 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实测值频率计数。表中的每个交叉分类构成一个单元格，每个分类变量称为一个因子。因变量为交叉制表单元格中的个案数（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极大似然参数。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格。单元结构变量允许您为不完整的表定义结构零、在模型中包括偏移项、拟合对数率模型或实施边际表的调整方法。对比变量允许计算广义对数优势比 (GLOR)。

模型信息和拟合优度统计自动显示。还可以显示各种统计和绘图，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。** 实测和期望的频率；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。

一般对数线性分析数据注意事项

**数据。** 因子是分类变量，单元格协变量是连续变量。当模型中有协变量时，会将单元格中个案的协变量平均值应用于该单元格。对比变量是连续的。它们用于计算广义对数几率比。对比变量的值是预期单元格计数的对数线性组合的系数。

单元格结构变量指定权重。例如，如果某些单元格为结构零，那么单元格结构变量的值为 0 或 1。请勿使用单元格结构变量来加权聚集数据。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：泊松和多项式。

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。
- 在单元格中出现观察的事件在统计上独立于其他单元格的单元格计数。

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。
- 单元格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。

获取一般对数线性分析

1. 从菜单中选择：

分析 > 对数线性 > 一般...

2. 在“常规对数线性分析”对话框中，选择最多 10 个因子变量。

(可选) 您可以执行以下操作：

- 选择单元格协变量。
- 选择单元格结构变量以定义结构零或包含偏移量项。
- 选择对比变量。

## 一般对数线性分析模型

**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。

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双阶交互。

### 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。

## 构建项和定制项

### 构建项：

当您想要为一组选定因子和协变量的所有组合包含特定类型（例如主效应）的非嵌套项时，使用该选项。

### 构建定制项

如果要包含嵌套项，或者想要按变量显式构建任何项，请使用该选项。构建嵌套项包括以下步骤：

## 一般对数线性分析：选项

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

**显示。** 多个统计可供显示：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（针对观察的单元格计数和期望的单元格计数的调整残差或偏差残差）。此外，还可以显示调整后残差或偏差残差的正态概率图和去趋势正态图。

**置信区间。** 可以调整参数估计值的置信区间。



**标准。** 使用牛顿-拉弗逊法来获取极大似然参数估算值。您可以为最大迭代次数、收敛准则和 Delta（为了获得初始近似值而对所有单元格添加的常量）输入新值。Delta 保留在饱和模型的单元格中。

## 一般对数线性分析：保存

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。

保存值指汇总数据（列联表中的单元格），即使数据记录在数据编辑器的各个实测值中。如果为未汇总的数据保存残差或预测值，那么将列联表中单元格的保存值输入到该单元格中每个案例的数据编辑器中。要使保存值有意义，应汇总数据以获得单元格计数。

可以保存四种残差：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。也称为简单或原始残差，它是观测单元格计数和期望单元格计数间的差值。
- 标准化残差。残差除以其标准误差的估计值。标准化残差也称作 Pearson 残差。
- 调整残差。标准化残差除以其估计的标准误差。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。似然比卡方统计量 (G 平方) 的单个添加项的带符号平方根，其中符号是残差的符号 (实测计数减去预期计数)。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：

- 计算观察到的单元格频率和期望单元格频率的线性组合，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）。

请参阅命令语法参考以获取完整的语法信息。

## 分对数的对数线性分析

“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、单元格协变量可以是连续变量，但它们不是针对每个案例应用的。单元格的加权协变量平均值应用于该单元格。因变量的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式 Logit 模型。此过程使用 Newton-Raphson 算法估算分对数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。单元结构变量允许您为不完整的表定义结构零、在模型中包括偏移项、拟合对数率模型或实施边际表的调整方法。对比变量允许计算广义对数优势比 (GLOR)。对比变量的值是预期单元格计数的对数线性组合的系数。

模型信息和拟合优度统计自动显示。还可以显示各种统计和绘图，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着它们的大小和它们生活的四个湖泊而变化？研究发现，较小的鳄鱼更喜欢爬行动物而不是鱼类的几率比较大的鳄鱼低 0.70 倍；此外，在第 3 号湖中，主要选择爬行动物而不是鱼类的几率最高。

**统计。** 观察到的和预期的频率；原始残差、调整残差和偏差残差；设计矩阵；参数估计；广义对数优势比；瓦尔德统计；和置信区间。图：调整残差、偏差残差和正态概率图。

Logit 对数线性分析数据注意事项

**数据。** 因变量是分类变量。因子是分类变量。单元格协变量可以是连续的，但当模型中有协变量时，单元格中案例的平均协变量值将应用于该单元格。对比变量是连续的。它们用于计算广义对数优势比 (GLOR)。对比变量的值是预期单元格计数的对数线性组合的系数。

单元格结构变量指定权重。例如，如果某些单元格为结构零，那么单元格结构变量的值为 0 或 1。请勿使用单元格结构变量来加权聚集数据。而应使用“数据”菜单中的“加权个案”。

**假设。** 假设解释变量类别的每个组合内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。
- 单元格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。当您想要分析观察到的计数与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

1. 从菜单中选择：

**分析 > 对数线性 > Logit...**

2. 在“Logit 对数线性分析”对话框中，选择一个或多个因变量。

3. 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。

(可选) 您可以执行以下操作：

- 选择单元格协变量。
- 选择单元格结构变量以定义结构零或包含偏移量项。
- 选择一个或多个对比变量。

## Logit 对数线性分析：模型

**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。

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双阶交互。

### 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。

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**，则还会将一个单元格项 (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。** 在定制模型中包括对应于该因变量的常量。

## 构建项和定制项

### 构建项：

当您想要为一组选定因子和协变量的所有组合包含特定类型（例如主效应）的非嵌套项时，使用该选项。

### 构建定制项

如果要包含嵌套项，或者想要按变量显式构建任何项，请使用该选项。构建嵌套项包括以下步骤：

## Logit 对数线性分析：选项

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

**显示。** 多个统计可供显示：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。** 用于定制模型的图包含两个散点图矩阵（针对观察到的单元格计数和期望单元格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。

**准则。** Newton-Raphson 方法可用于获取最大似然参数估计。可以为最大迭代次数、收敛准则和 delta（添加到所有初始近似值单元格的常数）输入新值。Delta 保留在饱和模型的单元格中。

## Logit 对数线性分析：保存

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。

保存值指汇总数据（列联表中的单元格），即使数据记录在数据编辑器的各个实测值中。如果为未汇总的数据保存残差或预测值，那么将列联表中单元格的保存值输入到该单元格中每个案例的数据编辑器中。要使保存值有意义，应汇总数据以获得单元格计数。

可以保存四种残差：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。也称为简单或原始残差，它是观测单元格计数和期望单元格计数间的差值。
- 标准化残差。残差除以其标准误差的估计值。标准化残差也称作 Pearson 残差。
- 调整残差。标准化残差除以其估计的标准误差。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。似然比卡方统计量 (G 平方) 的单个添加项的带符号平方根，其中符号是残差的符号 (实测计数减去预期计数)。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：

- 计算观察到的单元格频率和期望单元格频率的线性组合，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）。

请参阅命令语法参考以获取完整的语法信息。

## 寿命表

在许多情况下，您可能希望检查两个事件之间的时间分布，例如雇佣时间（从员工被雇佣到离职的时间）。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）。出现这种情况有多种原因：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终端事件的概率。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。您还可以绘制生存或风险函数，并直观地比较它们以获得更详细的信息。

**统计。** 每组在每个时间区间的期初记入数、期末离开数、历险数、终端事件数、终结比例、生存比例、累积生存比例（和标准误差）、概率密度（和标准误差）以及风险率（和标准误差）；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。

#### 寿命表数据比较

**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。** 所关注事件的概率应仅取决于初始事件之后的时间（假设绝对时间下的概率不变）。也就是说，在不同时间进入研究的病例（例如，在不同时间开始治疗的患者）应该表现相似。删失案例和未删失案例之间也不应该有系统性差异。例如，如果许多删失病例是病情更严重的患者，则您的结果可能有偏差。

**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。如果您的观测值数量较少，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测值。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。如果您的协变量在同一案例的不同时间点可能具有不同的值，请使用 Cox 回归和瞬态协变量。

#### 创建寿命表

1. 从菜单中选择：

**分析 > 生存 > 寿命表...**

2. 选择一个数值生存变量。

3. 指定要检查的时间区间。

4. 选择一个状态变量定义已发生终端事件的个案。

5. 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：定义范围

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

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，那么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。
- 日志生存。在对数刻度上显示累积生存函数。
- 危险。以线性比例显示累积风险函数。
- 密度。显示密度函数。
- 减一生存。用于根据线性尺度按照被一减的方式绘制生存函数的散点图。

比较第一个因子的水平。如果有一阶控制变量，那么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，那么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。

请参阅命令语法参考以获取完整的语法信息。

## 参数加速故障时间模型

参数加速失败时间 (AFT) 模型分析使用非循环生命时间数据调用参数生存模型过程。参数化生存模型假设生存时间遵循已知分布，此分析拟合加速失败时间模型，其模型效应与生存时间成正比。

### 获取参数化加速故障时间模型分析

1. 从菜单中选择：

分析 > 生存 > 参数加速故障时间 (AFT) 模型

2. 选择源变量。

#### 时间

##### 生存分析函数

表示生存时间持续时间的单个数字变量。

##### 开始/结束

表示 **开始时间** 和 **结束时间** 的数字变量。

#### 状态

用于确定下列其中一个状态设置的单个可选字符串或数字变量：

##### 故障/事件

将记录映射到故障/事件类别。字符串状态变量的缺省值为 F。

##### 右方设限

将记录映射到正确的审查类别。字符串状态变量的缺省值为 R。

##### 左方设限

将记录映射到左侧审查类别。字符串状态变量的缺省值为 L。

##### 时间间隔设限

将记录映射到时间间隔检查类别。仅适用于 **开始/结束**。字符串状态变量的缺省值为 I。

##### 未映射值处理

控制要将未映射的记录映射到哪个类别。要删除未能映射的记录，请选择 **从分析中排除这些记录**。

对于 **生存**，所有个案的缺省状态为 **故障/事件**。对于 **开始/结束**，缺省状态为 **Interval Censoring**。单击 **定义事件** 按钮以定义状态变量的事件。

#### 协变量

要作为协变量处理的一个或多个可选数字变量。请注意，变量不能同时由 **协变量** 和 **固定因子** 指定。

#### 固定因子

要作为因子处理的一个或多个可选变量。变量不能同时由 **固定因子** 和 **协变量** 指定。

#### 左截断

仅针对 **Survival** 的左截断的单个可选数字变量。

## 参数加速故障时间模型: 条件

### 条件

用于指定常规条件的可选面板。

### 置信区间

可选百分比，用于指定回归参数的置信区间的级别。它必须是介于 0 与 100 之间的单个数字值。缺省值为 95。

### 缺失值

用于控制如何处理用户缺失值的选项:

#### 排除用户缺失值和系统缺失值

将用户缺失值视为有效值。这是缺省值。

#### 将用户缺失值作为有效值处理

忽略用户缺失值指定，并将其视为有效值。

### 状态处理

仅适用于 **开始/结束**。用于控制如何处理具有不正确状态字段的记录的选项:

#### 废弃有冲突的记录

删除有冲突的记录。这是缺省设置。

#### 根据状态获取时间信息(O)

根据状态获取时间信息。

#### 根据时间信息派生状态(R)

根据时间信息更改状态。

## 参数加速故障时间模型: 模型

### 模式

用于指定模型选项和设置的可选面板。

### 生存时间的分布

用于指定生存时间分布的选项。

#### 威布尔

指定威布尔分布。这是缺省设置。

#### 指数

指定指数分布。

#### 对数正态分布(L)

指定对数正态分布。

#### 对数 Logistic

指定 log-logistic 分布。

### 协变量设置

指定协变量。

### 因子设置

指定因子变量。

### 截距初始值

用于指定截距项的初始值的选项。如果指定，那么它必须是单个数字值，并且不能是 0。

### 刻度参数的初始值

用于控制刻度参数设置的选项。

#### 相应 OLS 回归的标准误差(N)

使用相应的普通最小二乘回归的标准误差作为初始值。

#### 反转相应 OLS 回归的标准误差

使用标准误差的倒数。

### 用户提供的值(S)

如果指定了单个数字值，那么该值将用作初始值。如果指定此参数，那么它必须大于 0。

## 参数加速故障时间模型: 估算

### 估算

一个可选面板，用于指定用于控制加速故障时间模型和可选特征选择过程的估算的设置。

#### 交替方向方法或多箱 (ADMM)

##### 快速方式

应用乘数 (ADMM) 的快速交替方向方法。这是缺省值。

##### 传统

应用传统 ADMM 算法。

#### 应用 L-1 规则化(Z)

执行控制特征选择的过程。**惩罚参数** 字段指定控制规则化过程的惩罚参数。它必须是大于 0 的单个值。缺省设置为 0.001。

#### 模型收敛条件

##### 参数收敛

指定参数的收敛条件。它必须是属于  $[0, 1)$  的单个数字值。缺省设置为 0.000001。对于 **类型**，可以选择 **ABSOLUTE** 以将绝对收敛应用于内部优化，或者选择 **RELATIVE** 以将相对收敛应用于内部优化。可选的 **值** 字段指定关键字。

##### 目标函数收敛

指定目标函数的收敛条件。它必须是属于  $[0, 1)$  的单个数字值。缺省设置为 0，这不会应用收敛条件。对于 **类型**，可以选择 **ABSOLUTE** 以将绝对收敛应用于内部优化，或者选择 **RELATIVE** 以将相对收敛应用于内部优化。可选的 **值** 字段指定关键字。

##### Hessian 收敛性

指定 Hessian 矩阵的收敛条件。它必须是属于  $[0, 1)$  的单个数字值。缺省设置为 0，这不会应用收敛条件。对于 **类型**，可以选择 **ABSOLUTE** 以将绝对收敛应用于内部优化，或者选择 **RELATIVE** 以将相对收敛应用于内部优化。可选的 **值** 字段指定关键字。

#### 残差收敛条件

用于控制优化过程的选项。

##### 原始残差和对偶残差(B)

同时应用主和对偶残差收敛准则。这是缺省设置。

##### 仅原始残差

应用主残差收敛准则。

##### 仅对偶残差

应用对偶残差收敛准则。

#### 方法

用于指定估计方法的可选参数。

##### 自动

根据样本数据集自动选择方法。这是缺省值。**预测变量的阈值数** 字段指定预测变量数的阈值，并且必须是大于 1 的单个整数。缺省值为 1000。

##### 牛顿-拉弗逊(W)

应用 Newton-Raphson 方法。

##### L-BFGS(G)

应用有限内存 BFGS 算法。**更新** 字段指定由受限内存 BFGS 算法维护的过去更新数，并且必须是大于或等于 1 的单个整数。缺省值是 5。

#### 迭代

##### 最大迭代次数

指定最大迭代数。它必须是属于  $[1, 100]$  的单个整数。缺省设置为 20。

### 最大逐步二分次数

指定步骤二分法的最大数目。它必须是属于 [1, 20] 的单个整数。默认设置为 5。

### 最大行搜索数

指定行搜索的最大数目。它必须是属于 [1, 100] 的单个整数。缺省设置为 20。

### 迭代过程的绝对收敛

指定外部迭代过程的绝对收敛。它必须是属于 (0, 1) 的单个数字值。缺省设置为 0.0001。

### 迭代过程的相对收敛

指定外部迭代过程的相对收敛。它必须是属于 (0, 1) 的单个数字值。缺省设置为 0.01。

## 参数加速故障时间模型: 打印

### 打印

用于控制表输出的可选面板。

#### 因子编码详细信息(F)

如果选择此选项, 那么将显示并打印因子的编码详细信息。如果没有有效的因素, 那么将忽略该过程。

#### 分配给回归参数的初始值(I)

如果选择此选项, 那么将显示估算过程中使用的初始值。

#### 模型迭代历史记录(M)

如果选择此选项, 那么将显示生存分析的迭代历史记录。在 **步骤数** 字段中, 指定 1 到 99999999 之间的步骤数。缺省设置为 1。

#### 选择结果包含(T)

控制功能部件选择的详细信息的显示。

#### 选定变量和未选定变量

在表中显示选定变量和未选定变量。

#### 仅所选变量

仅显示所选变量。

#### 仅未选择的变量

仅显示未选择的变量。要显示的最大变量数 字段指定表中打印的变量的最大数目。默认设置为 30。

## 参数加速故障时间模型: 预测

### 预测

一个可选面板, 用于对预测统计信息进行评分并将其保存到活动数据集。

#### 评分的时间值

##### 由因变量定义的时间值(T)

根据为参数生存模型指定的时间变量对 **预测** 进行评分。

##### 定期(R)

根据未来时间值对 **预测** 进行评分。时间间隔 字段指定时间间隔, 并且必须是大于 0 的单个数字值。时间段数 字段指定时间段数, 并且必须是 2 到 100 之间的单个数字整数。

##### 持续时间

根据定义未来时间值的持续时间对 **预测** 进行评分。它必须是单个数字变量。

#### 预测

##### 生存分析函数

对预测生存统计信息进行评分并将其保存到活动数据集。缺省定制变量名称 (或根名称) 为 PredSurvival。

##### 风险

对预测的危险进行评分并将其保存到活动数据集。缺省定制变量名称 (或根名称) 为 PredHazard。



### 累积风险(Z)

对预测的累积危险进行评分并将其保存到活动数据集。缺省定制变量名称 (或根名称) 为 PredCumHazard。

### 条件生存(D)

对预测的条件生存统计信息进行评分并将其保存到活动数据集。缺省定制变量名称 (或根名称) 为 PredConditionalSurvival。如果未指定 PASTTIME, 那么将忽略该进程。需要 **过去生存时间** 值, 并指定用于评分的过去时间值。它必须是单个数字变量。

## 参数加速故障时间模型: 图

### 图

#### Cox-Snell 残差图

选择 **显示图** 以创建 Cox-Snell 残差图。在 **分箱分割点数** 字段中, 指定从 1 到 10000 的数字。缺省设置为 100。

#### 函数图

用于控制函数图的选项。

#### 类型

##### 生存分析函数

创建生存函数的图。

##### 风险

为危险函数创建图。

##### 密度

为密度函数创建图。

#### 要显示的点数

指定 1 到 200 之间的函数点数。缺省设置为 100。

#### 图的协变量值

可选, 用于指定用户提供的值并将其分配给预测变量。缺省情况下, 将在有效的每个协变量的 **平均值** 和有效的每个因子的类别频率上创建指定的图。如果指定, 将根据模式的设置创建指定的图。如果存在任何重复的变量, 那么首先指定的变量将被识别, 其余变量将被忽略。有效变量必须包含在模型效应中。对于协变量, 用户提供的值必须是数字。省略有效变量表示缺省情况下将对因子和协变量分别使用类别频率和 **平均值**。如果对变量指定了无效值, 那么将不会绘制所请求的模式。

#### 图的因子值

可选, 用于指定用户提供的值并将其分配给预测变量。如果存在任何重复的变量, 那么首先指定的变量将被识别, 其余变量将被忽略。有效变量必须包含在模型效应中。省略有效变量表示缺省情况下将对因子和协变量分别使用类别频率和平均值。如果对变量指定了无效值, 那么将不会绘制所请求的模式。

#### 针对下列各项绘制单独的线

用于指定用于绘制折线图的分类变量的选项。

#### 图表中的最大行数

如果指定了 **Separate lines for**, 那么指定图表中的最大行数。缺省设置为 10。

## 参数加速故障时间模型: 导出

### 导出

选择 **将模型信息导出到 XML 文件** 以将模型和参数信息写入 PMML 文件进行评分。必须指定要保存的 PMML 文件的目录和文件名。

## 生存 AFT 定义状态变量的事件

选定的一个或多个状态变量值的出现表明发生了针对这些情况的终端事件。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。

## 参数加速故障时间模型: 选择类别

"选择类别" 设置提供选项以选择表示要作为基线建模以进行比较的类别的值。

选择类别

单击 "最后一个类别" 以打开 "选择类别" 对话框。

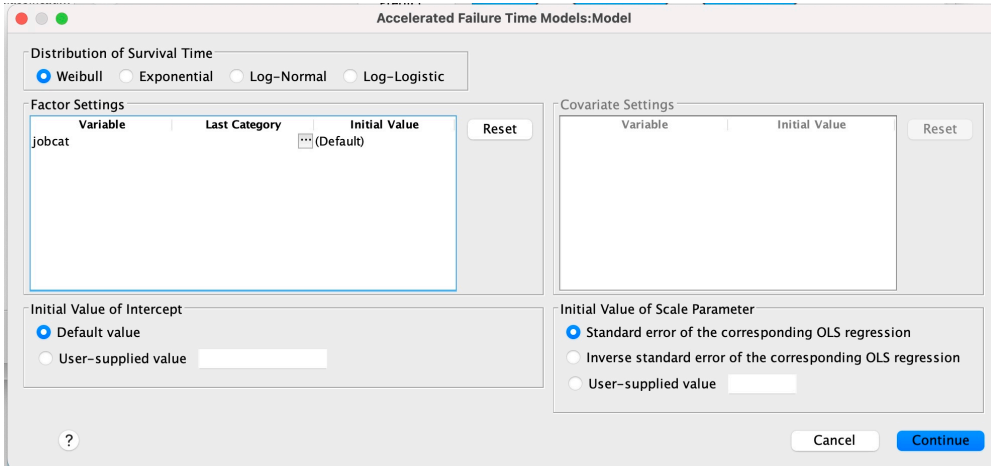


图 3: 加速生命周期模型-对话框-类别

要将类别指定为基线, 请从 "选择类别" 对话框中选择值。

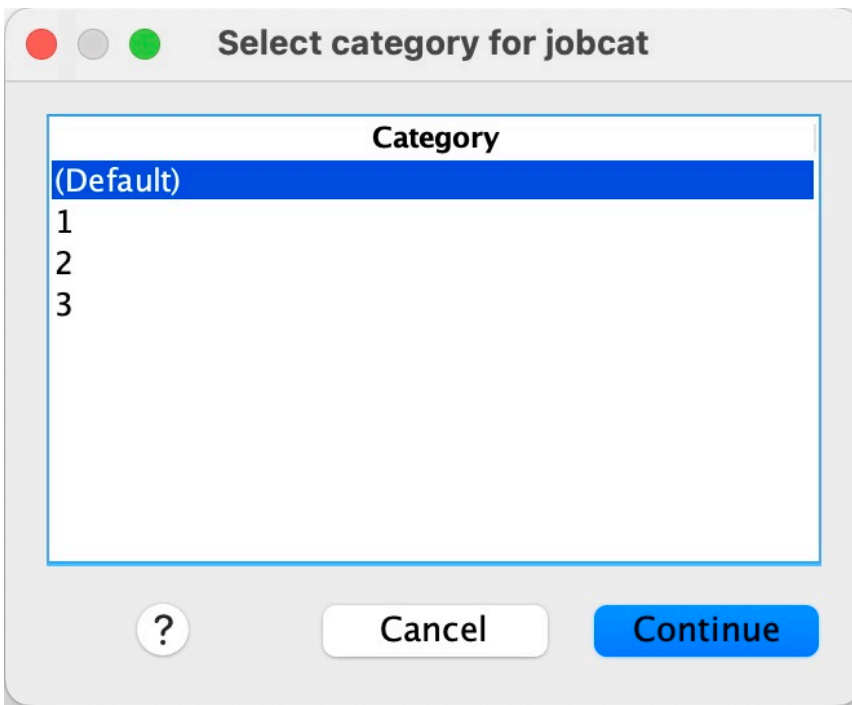


图 4: 加速生命周期模型-对话框-选择类别

单击 Continue (继续)。

## 参数化共享脆弱模型

参数共享的 FRailty 模型生存分析使用循环生命时间数据输入来启动参数生存模型过程。参数生存模型假定生存时间遵循已知分布, 并且此分析将一个脆弱的术语合并到参数生存模型中。它被视为一个随机组件, 以说明由于个体或群体级别的可变性而产生的未观察到的影响。

## 获取参数化共享脆弱模型分析

1. 从菜单中选择:

分析 > 生存 > 参数化共享脆弱模型

2. 选择源变量。

### 时间

#### 生存分析函数

生存时间由一个变量表示以表示结束时间。开始时间将设置为 0。

#### 开始/结束

表示 **开始时间** 和 **结束时间** 的数字变量。

### 主体

需要运行该过程。指定主题标识的单个变量。

### 时间间隔

指定用于标识共享同一主题标识的不同重复记录的区间号的单个和数字变量。

### 状态

用于确定下列其中一个状态设置的单个可选字符串或数字变量:

#### 故障/事件

将记录映射到故障/事件类别。字符串状态变量的缺省值为 F。

#### 右方设限

将记录映射到正确的审查类别。字符串状态变量的缺省值为 R。

#### 未映射值处理

控制要将未映射的记录映射到哪个类别。要删除未能映射的记录, 请选择 **从分析中排除这些记录**。

单击 **定义事件** 按钮以定义状态变量的事件。

### 协变量

要作为协变量处理的一个或多个可选数字变量。请注意, 变量不能同时由 **协变量** 和 **固定因子** 指定。

### 固定因子

要作为因子处理的一个或多个可选变量。变量不能同时由 **固定因子** 和 **协变量** 指定。

## 参数化共享脆弱模型: 条件

### 条件

用于指定常规条件的可选面板。

#### 置信区间

可选百分比, 用于指定回归参数的置信区间的级别。它必须是介于 0 与 100 之间的单个数字值。缺省值为 95。

#### 显著性水平

用于指定脆弱组件的似然比检验的显著性水平的选项。它必须是介于 0 和 1 之间的单个数字值。缺省设置为 0.05。

#### 缺失值

用于控制如何处理用户缺失值的选项:

##### 排除用户缺失值和系统缺失值

将用户缺失值视为有效值。这是缺省值。

##### 将用户缺失值作为有效值处理

忽略用户缺失值指定, 并将其视为有效值。

#### 区间处理

用于控制如何处理其时间间隔与开始时间和结束时间冲突的记录的选项。如果有两个时间变量在主对话框中指定了 Interval 变量, 那么它将生效。

### 废弃冲突的记录(D)

如果时间间隔值与开始时间和结束时间冲突，那么将废弃整个主题序列记录。这是缺省设置。

### 根据开始时间和结束时间发现时间间隔值(R)

从开始时间和结束时间发现时间间隔值。

## 参数化共享脆弱模型: 模型

### 模式

用于指定模型选项和设置的可选面板。

### 生存时间的分布

用于指定生存时间分布的选项。

#### 威布尔

指定威布尔分布。这是缺省设置。

#### 指数

指定指数分布。

#### 对数正态分布(L)

指定对数正态分布。

#### 对数 Logistic

指定 Log-logistic 分布。

### 协变量设置

指定协变量。

### 因子设置

指定因子变量。

### 截距初始值

用于指定截距项的初始值的选项。如果指定，那么它必须是单个数字值，并且不能是 0。

### 刻度参数的初始值

用于控制刻度参数设置的选项。

#### 相应 OLS 回归的标准误差(N)

使用相应的普通最小二乘回归的标准误差作为初始值。

#### 反转相应 OLS 回归的标准误差

使用标准误差的倒数。

#### 用户提供的值(S)

如果指定了单个数字值，那么该值将用作初始值。如果指定此参数，那么它必须大于 0。

### Frailty 组件

可选参数，用于指定 fRaily 组件的分布。

#### Gamma

指定伽玛分布。这是缺省设置。

#### 逆高斯(U)

指定逆高斯分布。

#### 初始方差值

指定 fRaily 组件的方差的初始值。它必须是大于 0 的单个数字值。对于伽玛分布，缺省值为 1.0，对于逆高斯分布，缺省值为 0.1。

## 参数化共享脆弱模型: 估算

### 估算

一个可选面板，用于指定用于控制共享脆弱模型估算和可选功能部件选择过程的设置。

## 交替方向方法或多箱 (ADMM)

### 快速方式

应用乘数 (ADMM) 的快速交替方向方法。这是缺省值。

### 传统

应用传统 ADMM 算法。

## 应用 L-1 规则化(Z)

执行控制特征选择的过程。**惩罚参数** 字段指定控制规则化过程的惩罚参数。它必须是大于 0 的单个值。缺省设置为 0.001。

## 模型收敛条件

### 参数收敛

指定参数的收敛条件。它必须是属于  $[0, 1)$  的单个数字值。缺省设置为 0.000001。对于 **类型**，可以选择 **ABSOLUTE** 以将绝对收敛应用于内部优化，或者选择 **RELATIVE** 以将相对收敛应用于内部优化。可选的 **值** 指定收敛类型的数字阈值。

### 目标函数收敛

指定目标函数的收敛条件。它必须是属于  $[0, 1)$  的单个数字值。缺省设置为 0，这不会应用收敛条件。对于 **类型**，可以选择 **ABSOLUTE** 以将绝对收敛应用于内部优化，或者选择 **RELATIVE** 以将相对收敛应用于内部优化。可选的 **值** 指定收敛类型的数字阈值。

### Hessian 收敛性

指定 Hessian 矩阵的收敛条件。它必须是属于  $[0, 1)$  的单个数字值。缺省设置为 0，这不会应用收敛条件。对于 **类型**，可以选择 **ABSOLUTE** 以将绝对收敛应用于内部优化，或者选择 **RELATIVE** 以将相对收敛应用于内部优化。可选的 **值** 指定收敛类型的数字阈值。

## 残差收敛条件

用于控制优化过程的选项。

### 原始残差和对偶残差(B)

同时应用主和对偶残差收敛准则。缺省情况下，此设置为 ""。

### 仅原始残差

应用主残差收敛准则。

### 仅对偶残差

应用对偶残差收敛准则。

## 方法

用于指定估计方法的可选参数。

### 自动

根据样本数据集自动选择方法。缺省情况下选择此方法。**预测变量的阈值数** 字段指定预测变量数的阈值，并且必须是大于 1 的单个整数。缺省值为 1000。

### 牛顿-拉弗逊(W)

应用 Newton-Raphson 方法。

### L-BFGS(G)

应用有限内存 BFGS 算法。**更新** 字段指定由有限内存 BFGS 算法维护的过去更新数，并且必须是大于或等于 1 的单个整数。缺省值是 5。

## 迭代

### 最大迭代次数

指定最大迭代数。它必须是属于  $[1, 300]$  的单个整数。缺省设置为 20。

### 最大逐步二分次数

指定步骤二分法的最大数目。它必须是属于  $[1, 200]$  的单个整数。默认设置为 5。

### 最大行搜索数

指定最大行搜索数。它必须是属于  $[1, 300]$  的单个整数。缺省设置为 20。

### 迭代过程的绝对收敛

指定外部迭代过程的绝对收敛。它必须是属于  $(0, 1)$  的单个数字值。缺省设置为 0.0001。

### 迭代过程的相对收敛

指定外部迭代过程的相对收敛。它必须是属于  $(0, 1)$  的单个数字值。缺省设置为 0.01。

## 参数化共享脆弱模型: 打印

### 打印

用于控制表输出的可选面板。

#### 因子编码详细信息(F)

如果选择此选项, 那么将显示并打印因子的编码详细信息。如果没有有效的因素, 那么将忽略该过程。

#### 分配给回归参数的初始值

如果选择此选项, 那么将显示估算过程中使用的初始值。

#### 模型迭代历史记录(M)

如果选择此选项, 那么将显示生存分析的迭代历史记录。在 **步骤数** 字段中, 指定 1 到 99999999 之间的步骤数。缺省设置为 1。

## 参数共享脆弱模型: 预测

### 预测

一个可选面板, 用于对预测统计信息进行评分并将其保存到活动数据集。

#### 评分的时间值

##### 由因变量定义的时间值(T)

根据为参数生存模型指定的时间变量对 **预测** 进行评分。

##### 定期(R)

根据未来时间值对 **预测** 进行评分。**时间间隔** 字段指定时间间隔, 并且必须是大于 0 的单个数字值。**时间段数** 字段指定时间段数, 并且必须是 2 到 100 之间的单个数字整数。

##### 持续时间

根据定义未来时间值的持续时间对 **预测** 进行评分。它必须是单个数字变量。

#### 预测

##### 生存分析函数

对预测生存统计信息进行评分并将其保存到活动数据集。缺省定制变量名称 (或根名称) 为 PredSurvival。

##### 风险

对预测的危险进行评分并将其保存到活动数据集。缺省定制变量名称 (或根名称) 为 PredHazard。

##### 累积风险(Z)

对预测的累积危险进行评分并将其保存到活动数据集。缺省定制变量名称 (或根名称) 为 PredCumHazard。

##### 条件生存(D)

对预测的条件生存统计信息进行评分并将其保存到活动数据集。缺省定制变量名称 (或根名称) 为 PredConditionalSurvival。如果未指定 PASTTIME, 那么将忽略该进程。需要 **过去生存时间** 值, 并指定用于评分的过去时间值。它必须是单个数字变量。

##### 无条件生存

对预测的无条件生存统计信息进行评分并将其保存到活动数据集。缺省情况下禁止该关键字。如果指定了此参数, 那么可以后跟在括号中指定的可选项用户提供的变量名 (或根名称)。缺省名称为 PredUnCondSurvival。

##### 无条件危险

对预测的无条件危险统计信息进行评分并将其保存到活动数据集。缺省情况下禁止该关键字。如果指定了此参数, 那么可以后跟在括号中指定的可选项用户提供的变量名 (或根名称)。缺省名称为 PredUncondHazard。

##### 无条件和危险

对预测的无条件累积危险统计信息进行评分并将其保存到活动数据集。缺省情况下禁止该关键字。如果指定了此参数, 那么它可能后跟用户提供的可选变量名 (或在括号中指定的根名称)。缺省名称为 PredUncondCumHazard。

## 参数化共享脆弱模型: 图

### 图

#### 函数图

用于控制函数图的选项。

#### 类型

##### 生存分析函数

为无条件生存函数创建图。

##### 风险

创建无条件危险函数的图。

##### 密度

为密度函数创建图。

#### 要显示的点数

指定 1 到 200 之间的函数点数。缺省设置为 100。

#### 图的协变量值

用于指定用户提供的值并将其分配给预测变量的选项。缺省情况下，将在有效的每个协变量的平均值处创建指定的图。如果指定了此选项，那么将根据模式的设置来创建指定的图。如果存在任何重复的变量，那么首先指定的变量将被识别，其余变量将被忽略。有效变量必须包含在模型效应中。对于协变量，用户提供的值必须是数字。省略有效变量表示缺省情况下将对协变量使用平均值。如果对变量指定了无效值，那么将不会绘制所请求的模式。

#### 图的因子值

用于指定用户提供的值并将其分配给预测变量的选项。缺省情况下，将在有效的每个因子的类别频率上创建指定的图。如果指定了此选项，那么将根据模式的设置来创建指定的图。如果存在任何重复的变量，那么首先指定的变量将被识别，其余变量将被忽略。有效变量必须包含在模型效应中。省略有效变量表示缺省情况下将对因子使用类别频率。如果对变量指定了无效值，那么将不会绘制所请求的模式。

#### 针对下列各项绘制单独的线

用于指定用于绘制折线图的分类变量的选项。

#### 图表中的最大行数

如果指定了 **Separate lines for**，那么指定图表中的最大行数。缺省设置为 10。

## 参数化共享脆弱模型: 导出

### 导出

选择 **将模型信息导出到 XML 文件** 以将模型和参数信息写入 PMML 文件进行评分。必须指定要保存的 PMML 文件的目录和文件名。

## 参数化共享脆弱模型: 定义事件

用于定义状态的选项。如果省略状态变量，那么失败或事件将成为所有案例的缺省状态。

#### 1. 从菜单中选择

分析 > 生存 > 参数化共享脆弱模型 ...

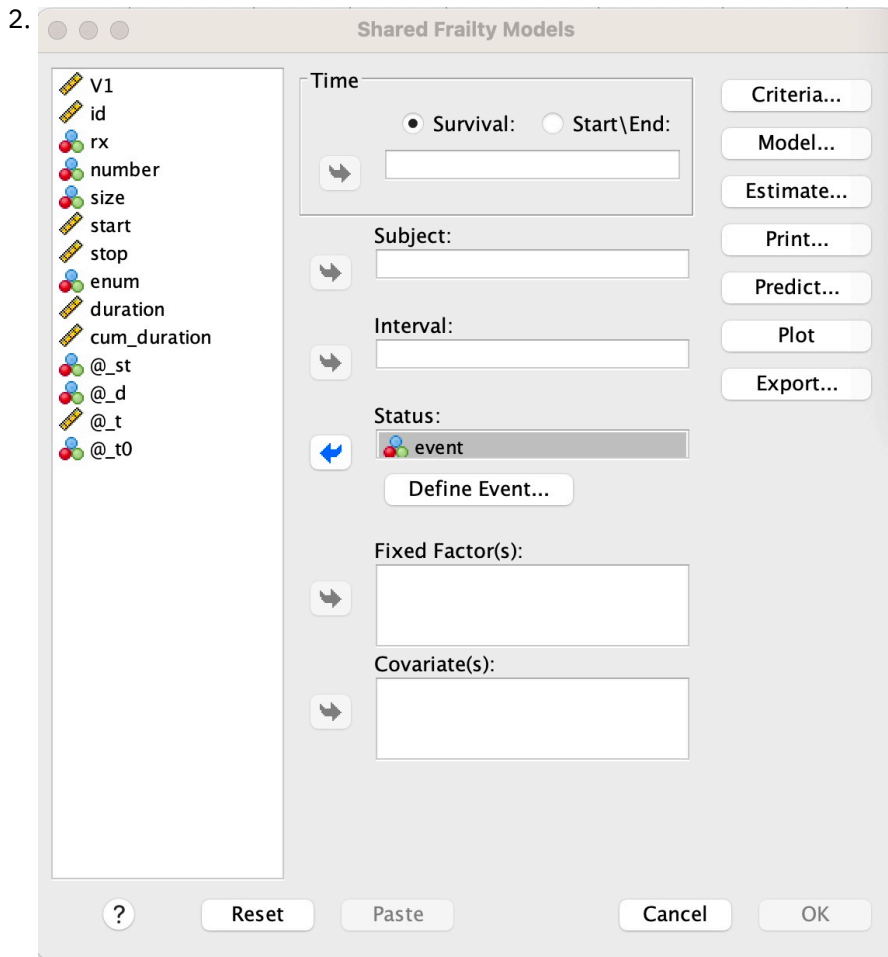


图 5: 共享残障模型-对话框-状态

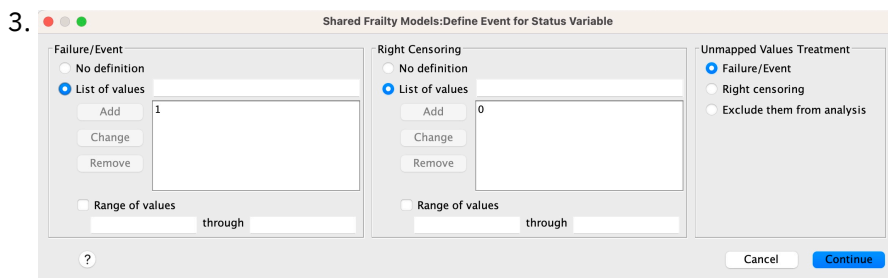


图 6: 共享残障模型-对话框-状态-定义事件

## 参数化共享脆弱模型-示例

### 示例 1

具有 x1 BY x2 的 SURVREG 循环 Y

/MODEL SUBJECT = ID FRAILTY=GAMMA DISTRIBUTION=WEIBULL。

在协变量 x1 和因子 x2 上拟合了参数共享-fRaily 生存模型 y。

生存时间由单个变量 y 表示。

主体由变量标识标识。

假设生存时间遵循威布尔分布。

假设 fRaily 的方差遵循 Gamma 分布。



在生存分析中使用所有有效记录。

## 示例 2

具有 x1 BY x2 的 SURVREG 循环 Y

```
/MODEL SUBJECT = ID FRAILITY=INV_GAIS DISTRIBUTION=LOG_NORMAL INTERVAL=z。
```

在协变量 x1 和因子 x2 上拟合了参数共享-fRailty 生存模型 y。

生存时间由表示开始时间和结束时间的两个变量 y1 和 y2 表示。

主体由变量标识标识。

假定生存时间遵循对数正态分布。

假设 fRailty 的方差遵循逆高斯分布。

时间间隔由变量 z 定义。对于每个主题，该过程仅使用不冲突的记录，并从分析中排除第一个失败状态之后的所有记录。

## 示例 3

具有 x1 BY x2(1 的 SURVREG 循环 y1 y2 )

```
/MODEL SUBJECT = 标识 FRAILITY=INV_GA 度分布 = LOG_LOGISTIC
```

```
/STATUS VARIABLE=event FAILURE=1 RIGHT=0。
```

在协变量 x1 和因子 x2 上拟合了参数共享-fRailty 生存模型 y。生存分析函数

时间分别表示两个变量 y1 和 y2 表示开始时间和结束时间。对于因子 x2，类别 "1" 指定为要建模的基线。

主体由变量标识标识。

假定生存时间遵循 Log-logistic 分布。

假设 fRailty 的方差遵循逆高斯分布。

指定变量事件以分别使用 1 和 0 表示故障和权限检查来定义状态。

## 示例 4

具有 x1 BY x2 的 SURVREG 循环 Y

```
/MODEL SUBJECT = 标识
```

```
/STATUS VARIABLE=event FAILURE=1 RIGHT=0
```

```
/预测不存在的不存在的不存在的不存在的不存在的不存在的不存在的不存在
```

```
/FUNCTIONPLOT SURVIVAL HAZARD PLOTBY (x2)。
```

在协变量 x1 和因子 x2 上拟合了参数共享-fRailty 生存模型 y。生存时间由单个变量 y 表示。

主体由变量标识标识。

将对无条件或基于人口的生存，危险和累积危险进行评分并将其保存到活动数据集。

无条件或基于人口的生存曲线和危险曲线由 x2 中的类别分隔。

## 示例 5

具有 x1 BY x2 的 SURVREG 循环 Y

```
/MODEL SUBJECT = ID FRAILITY=GAMMA DISTRIBUTION=WEIBULL
```

```
/STATUS VARIABLE=event FAILURE=1 RIGHT=0
```

```
/ESTIMATION HCONVERGE=1e-12(RELATIVE) PCONVERGE=0 FCONVERGE=0 SELECTFEATURES=TRUE  
PENALTY=0.01。
```

在协变量 x1 和因子 x2 上拟合了参数共享-fRailty 生存模型 y。生存时间由单个变量 y 表示。

主体由变量标识标识。

假设生存时间遵循威布尔分布。

假设 fRailty 的方差遵循 Gamma 分布。

收敛标准基于 Hessian 矩阵。它使用 1e-12 作为相对收敛。

该模型包含用于控制正则化过程的惩罚项。惩罚参数设置为 0.01。

## 示例 6

具有 x1 BY x2 的 SURVREG 循环 Y

```
/MODEL SUBJECT = 标识
```

```
/STATUS VARIABLE=infect FAILURE=1 RIGHT=0
```

```
/ESTIMATION MAXLINESEARCH=100 MAXITER=50 MAXSTEPHALVING=20。
```

在协变量 x1 和因子 x2 上拟合了参数共享-fRailty 生存模型 y。生存时间由单个变量 y 表示。

主体由变量标识标识。

该过程指定行搜索的最大数目为 100，迭代的最大数目为 50，步骤二分法的最大数目为 20。

## 参数化共享脆弱模型-经常性数据的案例研究

参数化共享脆弱模型-经常性数据的案例研究

用例名称-处理侧效应。

行为者 --- 公共卫生调查员和实践者。

前置条件-基于生存时间，副作用状态和要调整的预测变量可用的已清除数据集。

描述-公共健康调查人员 Patrick 正在调查包含 20 名参与者的数据样本。这些参与者是在一项研究中被招募的，研究的是一种可能由新治疗引起的轻度副作用。治疗设计者声称，在副作用方面，男性和女性之间不会有差异。帕特里克想评价这样一个假说。数据样本中包含的变量列示如下：

- patID: 用于标识唯一参与者的标识号。

- endTime: 遵循治疗的副作用生存时间 (以天为单位)，从治疗开始到报告副作用或检查 60 天内测量。

- sideEffect: 副作用状态，状态 = 0 (如果已审查) 和状态 = 1 (如果已报告轻微的副作用)。

-年龄: 参与人在研究期间的年龄。

-女性: 女性 = 0 (如果是男性)，女性 = 1 (如果是女性)。

可能会应用多个处理，这将生成针对特定参与者测量的重复时间的多个记录。对于每个记录，开始时间始终为 0，在数据样本中省略。帕特里克有兴趣将生存和危险功能可视化，通过控制其年龄和脆弱程度来绘制男性和女性之间的比较。他知道，对同一参与者进行的那些治疗是更相关的。通过假定生存时间遵循威布尔分布，Patrick 决定在 SPSS Statistics 中构建一个参数共享-脆弱生存模型，以考虑同一参与者的治疗依赖关系。

语法-

```
DATA LIST FREE  
/patID(F5.0) endTime(F5.0) sideEffect(F2.0) age(F5.2) female(F2.0) .  
BEGIN DATA .  
1 45 0 38.00 0  
2 26 1 20.00 1  
3 58 0 53.00 0  
4 31 1 37.00 1
```

```

4 24 0 37.00 1
4 50 0 37.00 1
5 20 1 51.00 0
5 38 1 51.00 0
6 30 0 35.00 1
7 22 1 58.00 1
8 53 1 29.00 1
8 49 1 29.00 1
9 25 0 45.00 0
9 25 0 45.00 0
10 27 0 33.00 1
11 34 1 21.00 1
11 40 0 21.00 1
11 49 0 21.00 1
12 42 1 26.00 0
13 25 0 40.00 0
14 21 1 52.00 0
14 32 1 52.00 0
15 56 0 28.00 1
15 34 0 28.00 1
16 30 0 41.00 0
16 29 0 41.00 0
17 25 1 27.00 0
18 26 1 54.00 1
18 36 1 54.00 1
19 27 0 39.00 0
20 58 1 22.00 1
20 54 0 22.00 1
20 43 1 22.00 1
END DATA.
SURVREG RECURRENT endTime WITH age BY female
/MODEL SUBJECT=patID FRAILTY=GAMMA DISTRIBUTION=WEIBULL
/ESTIMATION HCONVERGE=1e-12 PCONVERGE=0 FCONVERGE=0
/STATUS VARIABLE=sideEffect FAILURE=1 RIGHT=0
/FUNCTIONPLOT SURVIVAL HAZARD PLOTBY(female) .

```

### 概要:

由 Patrick 指定的语法将 `endTime` 指定为单个因时间变量。该过程自动假定每个记录的开始时间为 0。变量年龄和女性分别建模为协变量和因子。假定复发生存时间遵循威布尔分布。假定未观察到的脆弱项遵循伽玛分布，并对其方差成分进行建模。关于输出，“模型摘要”表提供了过程和模型信息。“案例处理摘要”表提供了故障/审查状态以及从分析中排除的那些案例的综合列表。

在 Patrick 的数据样本中，所有记录都是有效的，并包含在分析中。通过将似然与没有 `fRaiIty` 组件的对应模型的似然进行比较，共享 `fRaiIty` 模型无法达到显著水平 ( $p\text{-value} = 0.168$ )。Patrick 在想，是否有必要在模型中包含一个共享的-脆弱的术语。男性参与者的估计加速因子为 1.017，这是通过计算 [女性 = 0.0] 的估计回归系数 0.017 的指数获得的。其关联的 95% 置信区间为 (.688, 1.504)。这些结果表明，一个男性个体与同龄和脆弱的女性个体具有几乎相同的加速因子。在总体水平上，Patrick 分别为按年龄 (37.45 岁) 样本平均值进行评估的男性和女性绘制无条件生存曲线和危险曲线。

Patrick 确认，对于生存时间的任何固定值，平均一个男性和一个女性都应该具有相同的生存概率。有趣的是，尽管无条件危险图中显示了一个单峰形状，但帕特里克发现，在 60 天的时间内，人口危险实际上在增加。此行为可能暗示存在脆弱效应。为了进一步研究治疗所造成的副作用，Patrick 可以继续使用没有脆弱成分的模型，并比较男性和女性的行为。此外，他还可考虑与与会者进行为期超过 60 天的后续工作，以收集更多数据。

## Kaplan-Meier 生存分析

在许多情况下，您可能希望检查两个事件之间的时间分布，例如雇佣时间（从员工被雇佣到离职的时间）。但是，这种数据通常包含一些删失案例。删失案例是没有记录第二个事件的案例（例如，在研究结束时仍然为公司工作的员工）。Kaplan-Meier 过程是一种在存在删失案例的情况下估计事件时间模型的方法。Kaplan-Meier 模型基于估计事件发生时每个时间点的条件概率，并取这些概率的乘积极限来估计每个时间点的存活率。

**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。根据数据构建 Kaplan-Meier 模型可以让您比较两组之间的总生存率，以确定实验治疗是否优于传统治疗。您还可以绘制生存或风险函数，并直观地比较它们以获得更详细的信息。

**统计。** 生存表，包括时间、状态、累积生存和标准误差、累积事件和剩余数量；以及平均和中位生存时间，标准误差和 95% 置信区间。图：生存、风险、对数生存和 1 减生存。

## Kaplan-Meier 数据注意事项

**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。** 所关注事件的概率应仅取决于初始事件之后的时间（假设绝对时间下的概率不变）。也就是说，在不同时间进入研究的病例（例如，在不同时间开始治疗的患者）应该表现相似。删失案例和未删失案例之间也不应该有系统性差异。例如，如果许多删失病例是病情更严重的患者，则您的结果可能有偏差。

**相关过程。** Kaplan-Meier 过程使用一种计算寿命表的方法来估计每个事件发生时的生存或危险函数。寿命表过程使用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间隔，并且可能对处理大样本很有用。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。如果您的协变量在同一案例的不同时间点可能具有不同的值，请使用 Cox 回归和瞬态协变量。

获取 Kaplan-Meier 生存分析

1. 从菜单中选择：

**分析 > 生存 > Kaplan-Meier...**

2. 选择时间变量。

3. 选择一个状态变量标识已发生终端事件的个案。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。您还可以选择一个层变量，这将为变量的每个级别（层）生成单独的分析。

## Kaplan-Meier：为状态变量定义事件

输入表示已出现终端事件的值。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：比较因子级别

您可以请求统计以检验因子不同水平的生存分布的等同性。可用统计包括对数等级、Breslow 和 Tarone-Ware。选择其中一个选项以指定要进行比较：跨层整体检验、分层检验、跨层成对检验或分层成对检验。

**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级别的整体（而不是成对）比较。

- 对数等级。用于比较生存分布等同性的检验。在此检验中，所有时间点均赋予相同的权重。
- *Breslow*。用于比较生存分布等同性的检验。在每个时间点用带风险的个案数对时间点加权。
- *Tarone-Ware*。用于比较生存分布等同性的检验。在每个时间点用历险的个案数的平方根对时间点加权。
- 跨层整体检验。比较单个检验中的所有因子水平，以便检验生存曲线的相等性。
- 跨层成对检验。比较每个不同的因子级别对。成对比较趋势检验不可用。
- 分层检验。对每个层的所有因子级别的等同性执行单独的检验。如果您没有分层变量，那么不执行检验。
- 分层成对检验。比较每个层的每个不同的因子级别对。成对比较趋势检验不可用。如果您没有分层变量，那么不执行检验。

## Kaplan-Meier：保存新变量

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、生存函数的标准误差、风险函数和累积事件保存为新变量。

- 生存。累积生存概率估算值。缺省变量名为前缀 `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`。
- 生存标准误差。累积生存估计的标准误差。缺省变量名为前缀 `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。累积风险函数估算值。缺省变量名为前缀 `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。按生存时间和状态码将个案排序时事件的累积频率。缺省变量名为前缀 `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

您可以从 Kaplan-Meier 分析请求多种输出类型。

**统计。** 您可以选择为计算的生存函数显示的统计数据, 包括生存表、平均和中位数生存以及四分位数。如果已包括因子变量, 那么会针对每个组分别生成统计。

**图。** 绘图允许您直观地检查生存、减一生存、风险和对数生存函数。如果包含系数变量, 那么会为每组绘制函数。

- 生存。在线性刻度上显示累积生存函数。
- 减一生存。用于根据线性尺度按照被一减的方式绘制生存函数的散点图。
- 危险。以线性比例显示累积风险函数。
- 日志生存。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。

请参阅命令语法参考以获取完整的语法信息。

## Cox 回归分析

“Cox 回归”可以为时间-事件数据构建预测模型。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。从观测主项对生存函数的形状以及预测变量的回归系数进行评估; 然后将该模型应用于具有预测变量测量的新案例中。请注意, 删失主体 (即, 在实测期间未经历相关事件的主体) 的信息会有效地影响模型估算。

**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。

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

Cox 回归数据注意事项

**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。** 观测值应是独立的, 风险比应是时间恒定值; 即, 各个个案风险的比率不应随时间变化。后一个假设称为**比例风险假设**。

**相关过程。** 如果比例风险假设不成立 (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, 每个个案都出现终端事件), 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

1. 从菜单中选择:

**分析 > 生存 > Cox 回归...**

2. 选择时间变量。不分析时间值为负值的个案。

3. 选择一个状态变量, 然后单击**定义事件**。

4. 请选择一个或多个协变量。要包含交互项, 请选择交互中涉及的所有变量, 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

您可以详细指定 Cox 回归过程处理分类变量的方式。

**协变量。** 列出主对话框中在任何层指定的所有协变量，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。** 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。这种方法也称为 Helmert 反向对比。
- **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。假定类别间距相等。多项式对比仅适用于数字变量。
- **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**，那么可以选择**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

绘图可帮助您评估估计的模型和解释结果。可以绘制生存函数、风险函数、对数累积函数和一减生存函数。

### 图类型

#### 生存

在线性刻度上显示累积生存函数。

#### 危险

以线性比例显示累积风险函数。

#### 对数累积

向估计应用了  $\ln(-\ln)$  转换之后的累积生存估计。

#### 减一生存

用于根据线性尺度按照被一减的方式绘制生存函数的散点图。

### 协变量值的位置

可以将分类协变量移入**针对下列各项绘制单独的线条**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。

### 更改值

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。缺省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。

## Cox 回归：保存新变量

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差、对数累积估计、风险函数、偏残差和 DfBeta 以及线性预测变量  $X \cdot \text{Beta}$  保存为新变量。

### 生存函数

给定时间的累积生存函数值。该值等于生存到那个时间段概率。

### 生存函数的标准误差

累积生存估计的标准误差。

### 对数累积生存函数

向估计应用了  $\ln(-\ln)$  转换之后的累积生存估计。

### 风险函数

保存累积风险函数估计值（又称为 Cox-Snell 残差）。

### 偏残差

您可以根据生存时间来绘制部分残差，以检验比例风险假设。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

移除个案时系数中的估计变化。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。每个个案的以平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。

**注:** 如果正在使用依时协变量运行 Cox，那么仅保存 DfBeta。

### 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

您可以控制分析和输出的各个方面。

**模型统计。** 您可以获得模型参数的统计，包括  $\exp(B)$  的置信区间和估计值的相关性。可以在每个步骤或仅在最后一个步骤请求这些统计值。

**步进概率。** 如果选择了逐步法，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。** 允许您显示协变量平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，那么此显示不可用。

## Cox 回归：为状态变量定义事件

输入表示已出现终端事件的值。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。
- 指定附加迭代准则。
- 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。

## 计算依时协变量

在某些情况下，您希望计算 Cox 回归模型，但比例风险假设不成立。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。在此类情况下，您需要使用扩展的 Cox 回归模型，在该模型中可以指定 **依赖于时间的协变量**。

要分析此类模型，必须首先定义依赖于时间的协变量。为了便于执行此操作，提供了表示时间的“系统变量”。此变量称为  $T_$ 。您可以使用此变量来以两种常规方式定义时间相关协变量：

- 要测试与特定协变量相关的比例风险假设或估算允许非比例风险的扩展 Cox 回归模型，您需要将时间相关协变量定义为时间变量  $T_$  和相关协变量的函数。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。检验时间相关协变量的系数的显着性可知比例风险假设是否合理。
- 某些变量可以在不同时间段具有不同的值，但与时间没有系统地相关。在此类情况下，您需要定义 **分段的依赖于时间的协变量**，可以使用 **逻辑表达式** 来完成此操作。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。For example, if you have blood pressure that is measured once a week for the four weeks of your study (identified as  $BP1$  to  $BP4$ ), you can define the time-dependent covariate as  $(T_ < 1) * BP1 + (T_ \geq 1 \ \& \ T_ < 2) * BP2 + (T_ \geq 2 \ \& \ T_ < 3) * BP3 + (T_ \geq 3 \ \& \ T_ < 4) * BP4$ . 您会注意到，对于任何特定个案，括号中的其中一个术语正好等于 1，其余项都将等于 0。综上所述，此函数表示如果时间小于一周，请使用  $BP1$ ；如果时间大于一周但小于两周，请使用  $BP2$ ，依此类推。

在“计算依赖于时间的协变量”对话框中，可以使用函数构建控件来构建依赖于时间的协变量的表达式，也可以直接在“名称”文本区域中输入该表达式。注意，字符串常量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。任何生成的依赖于时间的协变量都需要作为协变量包含在 Cox 回归模型中。

## 计算依时协变量

1. 从菜单中选择  
    分析 > 生存 > Cox w/ Time-Dep Cov...
2. 输入依时协变量的表达式。
3. 单击**模型继续**“Cox 回归”。

注-确保在 Cox 回归模型中包含添加为协变量的新变量。

## 分类变量编码方案

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

与总平均值的偏差。在矩阵项中，对比具有以下形式：

```
mean ( 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

**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

```
mean (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)
```

## 赫尔默特

**Helmert 对比。** 比较自变量的类别与后续类别的平均值。一般矩阵格式是

```
mean (1/k  1/k  ...  1/k  1/k  1/k)
df(1) (  1 -1/(k-1)  ... -1/(k-1) -1/(k-1) -1/(k-1))
df(2) (  0   1  ... -1/(k-2) -1/(k-2) -1/(k-2))
      .
      .
df(k-2) (  0   0  ...   1  -1/2  -1/2)
df(k-1) (  0   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平均值。一般矩阵格式是

```
mean (  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/(k-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效应；第二自由度包含二次效应；第三自由度包含三次效应；对于更高阶的效应，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效应，而第二自由度包含二次效应。

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

```
mean (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，输入的第一行总是平均值（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** 的向量。

矩阵的其余行包含特殊对比，这些对比指示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) weights for mean calculation
(3 -1 -1 -1) compare 1st with 2nd through 4th
(0  2 -1 -1) compare 2nd with 3rd and 4th
(0  0  1 -1) compare 3rd with 4th
```

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 )
```

除平均值行外的每一行总和均为 0。每对不相交行的乘积总和也为 0：

```
Rows 2 and 3: (3)(0) + (-1)(2) + (-1)(-1) + (-1)(-1) = 0
Rows 2 and 4: (3)(0) + (-1)(0) + (-1)(1) + (-1)(-1) = 0
Rows 3 and 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

**指示符变量编码。**也称为哑元编码，在 LOGLINEAR 或 MANOVA 中不可用。新变量个数的代码为  $k-1$ 。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）之外的所有指示符变量都编码为 0。

## 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。

$$\begin{array}{cccc}
 (\sigma_1^2 & \sigma_2\sigma_1\rho_1 & \sigma_3\sigma_1\rho_1\rho_2 & \sigma_4\sigma_1\rho_1\rho_2\rho_3) \\
 (\sigma_2\sigma_1\rho_1 & \sigma_2^2 & \sigma_3\sigma_2\rho_2 & \sigma_4\sigma_2\rho_2\rho_3) \\
 (\sigma_3\sigma_1\rho_1\rho_2 & \sigma_3\sigma_2\rho_2 & \sigma_3^2 & \sigma_4\sigma_3\rho_3) \\
 (\sigma_4\sigma_1\rho_1\rho_2\rho_3 & \sigma_4\sigma_2\rho_2\rho_3 & \sigma_4\sigma_3\rho_3 & \sigma_4^2)
 \end{array}$$

**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等于  $\rho$ ，对于被第三个元素分隔的元素等于  $\rho^2$ ，依此类推，约束为  $-1 < \rho < 1$ 。

$$\begin{array}{cccc}
 (\sigma^2 & \sigma^2\rho & \sigma^2\rho^2 & \sigma^2\rho^3) \\
 (\sigma^2\rho & \sigma^2 & \sigma^2\rho & \sigma^2\rho^2) \\
 (\sigma^2\rho^2 & \sigma^2\rho & \sigma^2 & \sigma^2\rho) \\
 (\sigma^2\rho^3 & \sigma^2\rho^2 & \sigma^2\rho & \sigma^2)
 \end{array}$$

**AR(1)：异质。**这是具有异质方差的一阶自回归结构。任何两个元素之间的相关性，对于邻接元素等于  $r$ ，对于被另一个元素分开的两个元素是  $r^2$ ，依此类推。 $r$  约束在  $-1$  和  $1$  之间。

$(\sigma_1^2$	$\sigma_2\sigma_1\rho$	$\sigma_3\sigma_1\rho^2$	$\sigma_4\sigma_1\rho^3$ )
$(\sigma_2\sigma_1\rho$	$\sigma_2^2$	$\sigma_3\sigma_2\rho$	$\sigma_4\sigma_2\rho^2$ )
$(\sigma_3\sigma_1\rho^2$	$\sigma_3\sigma_2\rho$	$\sigma_3^2$	$\sigma_4\sigma_3\rho$ )
$(\sigma_4\sigma_1\rho^3$	$\sigma_4\sigma_2\rho^2$	$\sigma_4\sigma_3\rho$	$\sigma_4^2$ )

**ARMA(1,1)**。这是一阶自回归移动平均值结构。它包含齐次方差。两个元素之间的相关性对于相邻元素等于  $\rho$ ，对于被第三个元素分隔的元素等于  $\rho^2$ ，依此类推。自回归和移动平均参数的值限制在  $-1$  和  $1$  之间，包括  $-1$  和  $1$ 。

$(\sigma^2$	$\sigma^2\rho\rho$	$\sigma^2\rho\rho^2$	$\sigma^2\rho\rho^3$ )
$(\sigma^2\rho\rho$	$\sigma^2$	$\sigma^2\rho\rho$	$\sigma^2\rho\rho^2$ )
$(\sigma^2\rho\rho^2$	$\sigma^2\rho\rho$	$\sigma^2$	$\sigma^2\rho\rho$ )
$(\sigma^2\rho\rho^3$	$\sigma^2\rho\rho^2$	$\sigma^2\rho\rho$	$\sigma^2$ )

**复合对称**。此结构具有常数方差和常数协方差。

$(\sigma^2 + \sigma_1^2$	$\sigma_1$	$\sigma_1$	$\sigma_1$ )
$(\sigma_1$	$\sigma^2 + \sigma_1^2$	$\sigma_1$	$\sigma_1$ )
$(\sigma_1$	$\sigma_1$	$\sigma^2 + \sigma_1^2$	$\sigma_1$ )
$(\sigma_1$	$\sigma_1$	$\sigma_1$	$\sigma^2 + \sigma_1^2$ )

**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。

$(\sigma^2$	$\sigma^2\rho$	$\sigma^2\rho$	$\sigma^2\rho$ )
$(\sigma^2\rho$	$\sigma^2$	$\sigma^2\rho$	$\sigma^2\rho$ )
$(\sigma^2\rho$	$\sigma^2\rho$	$\sigma^2$	$\sigma^2\rho$ )
$(\sigma^2\rho$	$\sigma^2\rho$	$\sigma^2\rho$	$\sigma^2$ )

**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。

$(\sigma_1^2$	$\sigma_2\sigma_1\rho$	$\sigma_3\sigma_1\rho$	$\sigma_4\sigma_1\rho$ )
$(\sigma_2\sigma_1\rho$	$\sigma_2^2$	$\sigma_3\sigma_2\rho$	$\sigma_4\sigma_2\rho$ )
$(\sigma_3\sigma_1\rho$	$\sigma_3\sigma_2\rho$	$\sigma_3^2$	$\sigma_4\sigma_3\rho$ )
$(\sigma_4\sigma_1\rho$	$\sigma_4\sigma_2\rho$	$\sigma_4\sigma_3\rho$	$\sigma_4^2$ )

**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。

$(\sigma_1^2$	0	0	0)
(0	$\sigma_2^2$	0	0)
(0	0	$\sigma_3^2$	0)
(0	0	0	$\sigma_4^2$ )

**直积 AR1 (UN\_AR1)**。指定一个非结构化矩阵与另一个一阶自回归协方差矩阵的 Kronecker 乘积。第一个非结构化矩阵对多变量观测值进行建模，第二个一阶自回归协方差结构对时间或另一因子之间的数据协方差进行建模。

**非结构化直积 (UN\_UN)**。指定两个非结构化矩阵的 Kronecker 乘积，其中第一个矩阵对多变量观测值进行建模，第二个一阶矩阵对时间或另一因子之间的数据协方差进行建模。

**复合对称直积 (UN\_CS)**。指定一个非结构化矩阵与具有常数方差和协方差的另一个复合对称协方差矩阵的 Kronecker 乘积。第一个非结构化矩阵对多变量观测值进行建模，第二个复合对称协方差结构对时间或另一因子之间的数据协方差进行建模。

**因子分析：一阶**。此协方差结构具有元素间异质的项和元素间同质的项构成的异质方差。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。

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

**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。任何两个元素之间的协方差是其异质方差各项中第一项的乘积的平方根。

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

**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。

$$\begin{pmatrix} \sigma_1^2 & [\sigma_1^2 + \sigma_2^2]/2 - \lambda & [\sigma_1^2 + \sigma_3^2]/2 - \lambda & [\sigma_1^2 + \sigma_4^2]/2 - \lambda \\ ([\sigma_1^2 + \sigma_2^2]/2 - \lambda & \sigma_2^2 & [\sigma_2^2 + \sigma_3^2]/2 - \lambda & [\sigma_2^2 + \sigma_4^2]/2 - \lambda \\ ([\sigma_1^2 + \sigma_3^2]/2 - \lambda & [\sigma_2^2 + \sigma_3^2]/2 - \lambda & \sigma_3^2 & [\sigma_3^2 + \sigma_4^2]/2 - \lambda \\ ([\sigma_1^2 + \sigma_4^2]/2 - \lambda & [\sigma_2^2 + \sigma_4^2]/2 - \lambda & [\sigma_3^2 + \sigma_4^2]/2 - \lambda & \sigma_4^2 \end{pmatrix}$$

**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。

$$\begin{pmatrix} \sigma^2 & 0 & 0 & 0 \\ 0 & \sigma^2 & 0 & 0 \\ 0 & 0 & \sigma^2 & 0 \\ 0 & 0 & 0 & \sigma^2 \end{pmatrix}$$

**空间：幂**。此协方差结构在元素之间具有同质方差和异质相关性。 $d_{ij}$  是  $i^{\text{th}}$  和  $j^{\text{th}}$  测量之间的估计 Euclidean 距离。

$$\begin{pmatrix} \sigma^2 & \sigma^2 \rho^{d_{12}} & \sigma^2 \rho^{d_{13}} & \sigma^2 \rho^{d_{14}} \\ \sigma^2 \rho^{d_{12}} & \sigma^2 & \sigma^2 \rho^{d_{23}} & \sigma^2 \rho^{d_{24}} \\ \sigma^2 \rho^{d_{13}} & \sigma^2 \rho^{d_{23}} & \sigma^2 & \sigma^2 \rho^{d_{34}} \\ \sigma^2 \rho^{d_{14}} & \sigma^2 \rho^{d_{24}} & \sigma^2 \rho^{d_{34}} & \sigma^2 \end{pmatrix}$$

**空间：指数**。此协方差结构在元素之间具有同质方差和异质相关性。 $d_{ij}$  是  $i^{\text{th}}$  和  $j^{\text{th}}$  测量之间的估计 Euclidean 距离。

$$\begin{pmatrix} \sigma^2 & \sigma^2 \exp\{-d_{12}/\theta\} & \sigma^2 \exp\{-d_{13}/\theta\} & \sigma^2 \exp\{-d_{14}/\theta\} \\ \sigma^2 \exp\{-d_{12}/\theta\} & \sigma^2 & \sigma^2 \exp\{-d_{23}/\theta\} & \sigma^2 \exp\{-d_{24}/\theta\} \\ \sigma^2 \exp\{-d_{13}/\theta\} & \sigma^2 \exp\{-d_{23}/\theta\} & \sigma^2 & \sigma^2 \exp\{-d_{34}/\theta\} \\ \sigma^2 \exp\{-d_{14}/\theta\} & \sigma^2 \exp\{-d_{24}/\theta\} & \sigma^2 \exp\{-d_{34}/\theta\} & \sigma^2 \end{pmatrix}$$

**空间：高斯。** 此协方差结构在元素之间具有同质方差和异质相关性。  $d_{ij}$  是  $i^{\text{th}}$  和  $j^{\text{th}}$  测量之间的估计 Euclidean 距离。

$$\begin{pmatrix} \sigma^2 & \sigma^2 \exp\{-d_{12}/\rho^2\} & \sigma^2 \exp\{-d_{13}/\rho^2\} & \sigma^2 \exp\{-d_{14}/\rho^2\} \\ \sigma^2 \exp\{-d_{12}/\rho^2\} & \sigma^2 & \sigma^2 \exp\{-d_{23}/\rho^2\} & \sigma^2 \exp\{-d_{24}/\rho^2\} \\ \sigma^2 \exp\{-d_{13}/\rho^2\} & \sigma^2 \exp\{-d_{23}/\rho^2\} & \sigma^2 & \sigma^2 \exp\{-d_{34}/\rho^2\} \\ \sigma^2 \exp\{-d_{14}/\rho^2\} & \sigma^2 \exp\{-d_{24}/\rho^2\} & \sigma^2 \exp\{-d_{34}/\rho^2\} & \sigma^2 \end{pmatrix}$$

**空间：线性。** 此协方差结构在元素之间具有同质方差和异质相关性。  $d_{ij}$  是  $i^{\text{th}}$  和  $j^{\text{th}}$  测量之间的估计 Euclidean 距离，  $1_{ij}$  是一个指示符函数，如果  $\rho d_{ij} \leq 0$ ，则为 1，否则为 0。

$$\begin{pmatrix} \sigma^2 & \sigma^2(1 - \rho d_{12}) 1_{12} & \sigma^2(1 - \rho d_{13}) 1_{13} & \sigma^2(1 - \rho d_{14}) 1_{14} \\ \sigma^2(1 - \rho d_{12}) 1_{12} & \sigma^2 & \sigma^2(1 - \rho d_{23}) 1_{23} & \sigma^2(1 - \rho d_{24}) 1_{24} \\ \sigma^2(1 - \rho d_{13}) 1_{13} & \sigma^2(1 - \rho d_{23}) 1_{23} & \sigma^2 & \sigma^2(1 - \rho d_{34}) 1_{34} \\ \sigma^2(1 - \rho d_{14}) 1_{14} & \sigma^2(1 - \rho d_{24}) 1_{24} & \sigma^2(1 - \rho d_{34}) 1_{34} & \sigma^2 \end{pmatrix}$$

**空间：线性对数。** 此协方差结构在元素之间具有同质方差和异质相关性。  $d_{ij}$  是  $i^{\text{th}}$  和  $j^{\text{th}}$  测量之间的估计 Euclidean 距离，  $1_{ij}$  是指示符函数，如果  $\rho \log(d_{ij}) \leq 0$ ，则为 1，否则为 0。

$$\begin{pmatrix} \sigma^2 & \sigma^2(1 - \rho \log(d_{12})) 1_{12} & \sigma^2(1 - \rho \log(d_{13})) 1_{13} & \sigma^2(1 - \rho \log(d_{14})) 1_{14} \\ \sigma^2(1 - \rho \log(d_{12})) 1_{12} & \sigma^2 & \sigma^2(1 - \rho \log(d_{23})) 1_{23} & \sigma^2(1 - \rho \log(d_{24})) 1_{24} \\ \sigma^2(1 - \rho \log(d_{13})) 1_{13} & \sigma^2(1 - \rho \log(d_{23})) 1_{23} & \sigma^2 & \sigma^2(1 - \rho \log(d_{34})) 1_{34} \\ \sigma^2(1 - \rho \log(d_{14})) 1_{14} & \sigma^2(1 - \rho \log(d_{24})) 1_{24} & \sigma^2(1 - \rho \log(d_{34})) 1_{34} & \sigma^2 \end{pmatrix}$$

**空间：球形。** 此协方差结构在元素之间具有同质方差和异质相关性。  $r_{ij} = d_{ij}/\rho$ ，其中  $d_{ij}$  是  $i^{\text{th}}$  和  $j^{\text{th}}$  测量之间的估计 Euclidean 距离。  $1_{ij}$  是一个指示符函数，如果  $d_{ij} \leq \rho$ ，则为 1，否则为 0。

$$\begin{pmatrix} \sigma^2 & \sigma^2(1 - 3/2r_{12} + 1/2r_{12}^3) 1_{12} & \sigma^2(1 - 3/2r_{13} + 1/2r_{13}^3) 1_{13} & \sigma^2(1 - 3/2r_{14} + 1/2r_{14}^3) 1_{14} \\ \sigma^2(1 - 3/2r_{12} + 1/2r_{12}^3) 1_{12} & \sigma^2 & \sigma^2(1 - 3/2r_{23} + 1/2r_{23}^3) 1_{23} & \sigma^2(1 - 3/2r_{24} + 1/2r_{24}^3) 1_{24} \\ \sigma^2(1 - 3/2r_{13} + 1/2r_{13}^3) 1_{13} & \sigma^2(1 - 3/2r_{23} + 1/2r_{23}^3) 1_{23} & \sigma^2 & \sigma^2(1 - 3/2r_{34} + 1/2r_{34}^3) 1_{34} \\ \sigma^2(1 - 3/2r_{14} + 1/2r_{14}^3) 1_{14} & \sigma^2(1 - 3/2r_{24} + 1/2r_{24}^3) 1_{24} & \sigma^2(1 - 3/2r_{34} + 1/2r_{34}^3) 1_{34} & \sigma^2 \end{pmatrix}$$

**Toeplitz。** 此协方差结构的元素之间具有同质方差和异质相关性。 邻近元素间的相关在所有邻近元素对中具有齐性。 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。

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

**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。 邻近元素间的相关在所有邻近元素对中具有齐性。 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。

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

$(\sigma_4\sigma_1\rho_{31})$	$\sigma_4\sigma_2\rho_{21}$	$\sigma_4\sigma_3\rho_{31}$	$\sigma_4^2$
$(\sigma_1^2)$	$\sigma_2^2$	$\sigma_3^2$	$\sigma_4^2$
$(\sigma_2^2)$	$\sigma_2^2$	$\sigma_3^2$	$\sigma_4^2$
$(\sigma_3^2)$	$\sigma_3^2$	$\sigma_3^2$	$\sigma_4^2$
$(\sigma_4^2)$	$\sigma_4^2$	$\sigma_4^2$	$\sigma_4^2$

未结构化。这是一个非常一般的协方差矩阵。

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。

$(\sigma_1^2)$	$\sigma_2\sigma_1\rho_{21}$	$\sigma_3\sigma_1\rho_{31}$	$\sigma_4\sigma_1\rho_{41}$
$(\sigma_2\sigma_1\rho_{21})$	$\sigma_2^2$	$\sigma_3\sigma_2\rho_{32}$	$\sigma_4\sigma_2\rho_{42}$
$(\sigma_3\sigma_1\rho_{31})$	$\sigma_3\sigma_2\rho_{32}$	$\sigma_3^2$	$\sigma_4\sigma_3\rho_{43}$
$(\sigma_4\sigma_1\rho_{41})$	$\sigma_4\sigma_2\rho_{42}$	$\sigma_4\sigma_3\rho_{43}$	$\sigma_4^2$

方差成分。此结构为每个指定的随机效应分配一个标度恒等 (ID) 结构。

## Storing transaction logs in a database for high availability

You can configure transaction logs to be stored in a database where you can implement automatic replication and simpler disaster recovery.

### 开始之前

For more information about this feature and why it can be useful, read [Configure the transaction log location and the compensation log location for each server in the cluster before you enable high availability](#).

### 关于此任务

This solution uses two data centers. One is the primary data center and the other is a standby data center. The installation and configuration data from the primary data center is copied to the standby data center. Database replication is used to synchronize replication of runtime data from the primary database to the standby database.

### 过程

1. At the primary site, install IBM BPM on all nodes and create a deployment environment as you would normally.
2. Set up and configure the database.
  - a) Install the database for the primary data center.
  - b) Use generated database scripts to create database objects.
  - c) Install the database in corresponding standby data center.
  - d) Configure the databases to implement data replication between the primary database and the standby database.
3. Configure the transaction service.
  - a) Start the deployment environment at the primary data center.
  - b) For each cluster member in the application cluster and in the support cluster, create a data source on the cluster level. Configure the logs for the transaction service and the compensation service for each cluster member into the database. Use a unique prefix for each member. See detailed instructions in [Storing transaction and compensation logs in a relational database for high availability](#) in the WebSphere Application Server documentation.

4. When all cluster members in application and support clusters are configured, enable transaction high availability for each application and support cluster. On the **Configuration** page under **General Properties**, select **Enable failover of transaction log recovery**.

5. Restart the whole environment and make sure that there are no exceptions in the system log files.

When you start a process server that is configured to store transaction and compensation logs in a database, the transaction service can time out while the service waits for the data source to become available. If that happens, you see this error message:

```
WSVR0009E: Error occurred during startup com.ibm.ws.exception.RuntimeError:
com.ibm.ws.recoverylog.spi.InternalLogException: Failed to locate data source,
com.ibm.ws.recoverylog.spi.InternalLogException: Failed to locate data source at
com.ibm.ws.tx.util.WASTMHelper.asynchRecoveryProcessingComplete(WASTMHelper.java:176) at
com.ibm.tx.util.TMHelper.asynchRecoveryProcessingComplete(TMHelper.java:57)
```

If you encounter an error of this sort, increase the timeout value.

- a) Open the administrative console.
- b) Select **Servers > Application servers > server name**.
- c) Under **Server infrastructure**, select **Java and Process Management > Process Definition**.
- d) Under **Additional properties**, select **Java virtual machine > Custom properties > > New**.
- e) In the **Name** entry field, type `com.ibm.ws.recoverylog.custom.jdbc.impl.ConfigOfDataSourceTimeout`.
- f) In the **Value** entry field, set an integer timeout variable such as 30000 to represent a 30-second timeout. The timeout period is measured in milliseconds.
- g) Select **OK**.

## 下一步做什么

Devise a high availability and disaster recovery test plan that is appropriate for the business needs of your organization. The plan might include simulating a WebSphere Application Server ND failover and a cross-database failover to ensure that your system provides adequate business continuity.

## 贝叶斯单一样本推论：正态

此功能需要 SPSS Statistics Standard Edition 或“高级统计”选项。

“贝叶斯单样本推论：正态”过程提供用于通过描述后验分布特征对单样本和双样本配对 T 检验进行贝叶斯推论的选项。当具有正态数据时，可以使用正态先验来获取正态后验。

1. 从菜单中选择：

**分析 > 贝叶斯统计 > 单样本正态**

2. 从**可用变量**列表中选择相应的**检验变量**。应至少选择一个变量。

**注：**可用变量列表提供除日期和字符串变量以外的所有变量。

3. 选择所需的**贝叶斯分析**：

- **描述后验分布特征：**选中该选项时，将通过描述后验分布的特征所得出的视角做出贝叶斯推论。您可以通过集成其他冗余参数来调查感兴趣参数的边际后验分布，然后进一步构造可信区间来得出直接推论。这是缺省设置。
- **估算贝叶斯因子：**如果选中该选项，那么在比较原假设和替代假设之间的边际似然估计时，估算贝叶斯因子（贝叶斯推论中的一种著名的方法）占据一定的自然比。

贝叶斯因子	证据类别	贝叶斯因子	证据类别	贝叶斯因子	证据类别
>100	H1 极端证据	1-3	H1 轶事证据	1/30-1/10	H0 有力证据
30-100	H1 超有力证据	1	无证据	1/100-1/30	H0 超有力证据
10-30	H1 有力证据	1/3-1	H0 轶事证据	1/100	H0 极端证据



表 2: 常用于定义证据显著性的阈值 (继续)					
贝叶斯因子	证据类别	贝叶斯因子	证据类别	贝叶斯因子	证据类别
3-10	H1 适度证据	1/10-1/3	H0 适度证据		

$H_0$ : 原假设

$H_1$ : 其他假设

1

2

- **使用两种方法:** 如果选中该选项, 那么将同时使用描述后验分布特征和估算贝叶斯因子推论方法。
4. 选择和/或输入相应的数据方差和假设值设置。该表反映的是当前位于检验变量列表中的变量。随着检验变量列表中变量的添加或移除, 该表自动在其变量列中添加或移除相同的变量。

- 当检验变量列表中有一个或多个变量时, 将启用变量已知和方差值列。

#### 方差已知

当方差为已知时, 针对每个变量选择此选项。

#### 方差值

可选参数, 用于指定观察到的数据的方差值 (如果已知)。

- 当检验变量列表中有一个或多个变量, 且未选择描述后验分布特征时, 将启用原检验值和  $g$  值列。

#### 空检验值

必需参数, 用于在贝叶斯因子估算中指定空值。仅允许一个值, 缺省值为 0。

#### $g$ 值

指定该值以在贝叶斯因子估算中定义  $\psi^2 = g\sigma^2_x$ 。指定方差值时,  $g$  值缺省为 1。如果未指定方差值, 那么可以指定固定  $g$  或省略将其集成的值。

5. 您可以选择性地单击标准来指定第 83 页的『贝叶斯单一样本推论: 标准』设置 (可信区间百分比、缺失值选项和数值方法设置), 或者单击先验来指定第 84 页的『贝叶斯单一样本推论: 正态先验』设置 (先验类型, 例如推论参数、给定平均值方差或精度)。

## 贝叶斯单一样本推论: 标准

您可以为“贝叶斯单一样本推论”指定下列分析标准:

### 可信区间百分比

指定用于计算可信区间的显著性水平。缺省级别为 95%。

### 缺失值

指定用于控制缺失值的方法。

### 成对排除案例

这是缺省设置, 用于按分析排除具有缺失值的记录。检验中将省略包含该特定检验所用字段的缺失值的记录。

### 成列排除案例

该设置按列表排除包含缺失值的记录。如果记录包含在任何子命令上指定的任何字段的缺失值, 那么会从所有分析中排除此类记录。

**注:** 仅当为贝叶斯分析选择估算贝叶斯因子或使用两种方法选项时, 以下选项才可用。

### 数值方法

指定用于估算积分的数值方法。

### 自适应 Gauss-Lobatto 正交

这是缺省设置, 调用自适应 Gauss-Lobatto 正交方法。

<sup>1</sup> Lee, M.D., and Wagenmakers, E.-J. 2013. *Bayesian Modeling for Cognitive Science: A Practical Course*. Cambridge University Press.

<sup>2</sup> Jeffreys, H. 1961. *Theory of probability*. Oxford University Press.

### 容差

指定数值方法的容差值。缺省设置为 0.000001。仅当选择了自适应 Gauss-Lobatto 正交设置时，该选项才可用。

### 最大迭代数

指定自适应 Gauss-Lobatto 正交方法迭代的最大数量。此值必须是正整数。缺省设置为 2000。仅当选择了自适应 Gauss-Lobatto 正交设置时，该选项才可用。

### Monte Carlo 近似值

该选项调用 Monte Carlo 近似值方法。

### 设置定制种子

选中该选项时，可以在种子值字段中指定定制种子值。

### 种子数

为 Monte Carlo 近似值方法指定随机种子集。此值必须是正整数。缺省情况下，分配随机种子值。

### 蒙特卡洛样本的数量

指定为 Monte Carlo 近似值抽样的点数。此值必须是正整数。缺省值为 1000000。仅当选择 Monte Carlo 近似值设置时，该选项才可用。

## 贝叶斯单一样本推论：正态先验

您可以为“贝叶斯单一样本推论”指定下列先验分布标准：

注：很多应用研究员可能质疑指定先验的需求。参考先验可最大限度减少因为数据增加造成先验不堪重负的问题。在指定信息先验信息时，贝叶斯方法可有效地使用信息。指定先验的需要不应被视为妨碍使用贝叶斯分析。

### 先验方差/精度

提供用于定义方差和精度值的选项。

#### 方差

选择以指定方差参数的先验分布。选择此选项时，先验分布列表提供以下选项：

注：当已为某些变量指定数据方差时，将针对这些变量忽略以下设置。

- 扩散 - 缺省设置。指定扩散先验。
- 逆卡方 - 指定逆  $\chi^2(\nu_0, \sigma_0^2)$  的分布和参数，其中  $\nu_0 > 0$  是自由度， $\sigma_0^2 > 0$  是刻度参数。
- 逆伽玛 - 指定逆伽玛  $(\alpha_0, \beta_0)$  的分布和参数，其中  $\alpha_0 > 0$  是形状参数， $\beta_0 > 0$  是标度参数。
- Jeffreys S2 - 指定不提供信息的先前版本  $1/\sigma_0^2$ 。
- 0.
- Jeffreys S4 - 指定不提供信息的先前版本  $\propto 1/\sigma_0^4$ 。

#### 精度

选择以指定精度参数的先验分布。选择此选项时，先验分布列表提供以下选项：

- 伽玛 - 指定伽玛  $(\alpha_0, \beta_0)$  的分布和参数，其中  $\alpha_0 > 0$  是形状参数， $\beta_0 > 0$  是刻度参数。
- 卡方 - 指定  $\chi^2(\nu_0)$  的分布和参数，其中  $\nu_0 > 0$  是自由度。

#### 形状参数

指定逆伽玛分布的形状参数  $a_0$ 。必须输入一个大于 0 的值。

#### 标度参数

指定逆伽玛分布的刻度参数  $b_0$ 。必须输入一个大于 0 的值。刻度参数越大，分布越分散。

### 平均值给定方差/精度先验

指定视方差或精度参数而定的平均值参数的先验分布。

#### 正态

指定基于方差的正态  $(\mu_0, K^{-1}_0\sigma_0^2)$  或基于精度的正态  $(\mu_0, K_0/\sigma_0^2)$  的分布和参数，其中  $\mu_0 \in (-\infty, \infty)$  和  $\sigma^2 > 0$ 。

### 位置参数

输入用于指定分布的位置参数的数字值。

### 标度参数

指定逆伽玛分布的刻度参数  $b_0$ 。必须输入一个大于 0 的值。

### Kappa(K)

指定正态 ( $\mu_0, K^{-1}\sigma^2_0$ ) 或正态 ( $\mu_0, K_0/\sigma^2_0$ ) 中  $K_0$  的值。必须输入一个大于 0 的值（缺省值为 1）。

### 扩散

缺省设置，用于指定扩散先验  $\propto 1$ 。

## 贝叶斯单一样本推论：二项式

此功能需要 SPSS Statistics Standard Edition 或“高级统计”选项。

“贝叶斯单一样本推论：二项”过程提供用于对二项分布执行贝叶斯单一样本推论的选项。有关参数为  $\pi$ ，它表示可能导致成功或失败的固定数量的试验中的成功概率。请注意，每个试验相互独立，并且概率  $\pi$  在每个试验中保持相同。二项随机变量可被视为固定数量的独立 Bernoulli 试验的总和。

虽然并非必要，但是在估算二项参数时通常会选择 Beta 分布族中的先验。Beta 族是二项分布族的共轭，因此导致封闭式的后验分布仍属于 Beta 分布族。

1. 从菜单中选择：

分析 > 贝叶斯统计 > 单样本二项式

2. 从可用变量列表中选择相应的检验变量。应至少选择一个变量。

注：可用变量列表提供除日期和字符串变量以外的所有变量。

3. 选择所需的贝叶斯分析：

- **描述后验分布特征：**选中该选项时，将通过描述后验分布的特征所得出的视角做出贝叶斯推论。您可以通过集成其他冗余参数来调查感兴趣参数的边际后验分布，然后进一步构造可信区间来得出直接推论。这是缺省设置。
- **估算贝叶斯因子：**如果选中该选项，那么在比较原假设和替代假设之间的边际似然估计时，估算贝叶斯因子（贝叶斯推论中的一种著名的方法）占据一定的自然比。

表 3: 常用于定义证据显著性的阈值

贝叶斯因子	证据类别	贝叶斯因子	证据类别	贝叶斯因子	证据类别
>100	H1 极端证据	1-3	H1 轶事证据	1/30-1/10	H0 有力证据
30-100	H1 超有力证据	1	无证据	1/100-1/30	H0 超有力证据
10-30	H1 有力证据	1/3-1	H0 轶事证据	1/100	H0 极端证据
3-10	H1 适度证据	1/10-1/3	H0 适度证据		

$H_0$ : 原假设

$H_1$ : 其他假设

3

4

- **使用两种方法：**如果选中该选项，那么将同时使用描述后验分布特征和估算贝叶斯因子推论方法。

4. 选择和/或输入相应的成功类别和假设值设置。该表反映的是当前位于检验变量列表中的变量。随着检验变量中变量的添加或删除，该表会自动在其变量对列中添加或删除相同的变量。

<sup>3</sup> Lee, M.D., and Wagenmakers, E.-J. 2013. *Bayesian Modeling for Cognitive Science: A Practical Course*. Cambridge University Press.

<sup>4</sup> Jeffreys, H. 1961. *Theory of probability*. Oxford University Press.

- 当选择描述后验分布特征作为贝叶斯分析时，将启用成功类别列。
- 当选择估算贝叶斯因子或使用两种方法作为贝叶斯分析时，将启用所有可编辑列。

#### 原点

启用和禁用原比例选项。启用此设置后，原先验形状和原先验刻度选项都将禁用。

#### 原先验形状

在二项推论的原假设下指定形状参数  $a_0$ 。

#### 原先验刻度

在二项推论的原假设下指定刻度参数  $b_0$ 。

#### 原比例

在共轭先验分布的原假设下（以适应 Beta 和 Haldane 先验），指定形状参数  $a_0$  和刻度参数  $b_0$ 。有效范围是介于 0 与 1 之间的数字值。

#### 备用先验形状

必需参数，用于在二项推论的备用假设下指定  $a_0$ （如果要估算贝叶斯因子）。

#### 备用先验刻度

必需参数，用于在二项推论的备用假设下指定  $b_0$ （如果要估算贝叶斯因子）。

#### 成功类别

提供用于定义共轭先验分布的选项。所提供的选项指定在对检验值检验数据值时如何针对数字和字符串变量定义成功。

#### 最后一个类别

缺省设置，用于在按升序排序后使用类别中找到的最后一个数字值执行二项检验。

#### 第一个类别(F)

在按升序排序后使用类别中找到的第一个数字值执行二项检验。

#### 中点

使用大于或等于中点的数字值作为个案。中点值是最小和最大样本数据的平均值。

#### 分割点

使用大于或等于指定分界值的数字值作为个案。设置必须是单个数字值。

#### 级别

将用户指定的字符串值（可以大于 1）作为个案。使用逗号分隔不同的值。

5. 您可以选择性地单击**标准**来指定第 83 页的『贝叶斯单一样本推论：标准』设置（可信区间百分比、缺失值选项和数值方法设置），或者单击**先验**来指定第 86 页的『贝叶斯单一样本推论：二项式/泊松先验』设置（共轭或定制先验分布）。

## 贝叶斯单一样本推论：二项式/泊松先验

您可以为“贝叶斯单一样本推论”指定下列先验分布标准：

**注：**很多应用研究员可能质疑指定先验的需求。参考先验可最大限度减少因为数据增加造成先验不堪重负的问题。在指定信息先验信息时，贝叶斯方法可有效地使用信息。指定先验的需要不应被视为妨碍使用贝叶斯分析。

#### 形状参数

对于二项式先验，指定 Beta 分布的形状参数  $a_0$ 。

对于泊松先验，指定伽玛分布的形状参数  $a_0$ 。

必须输入一个大于 0 的值。

#### 标度参数

对于二项式先验，指定 Beta 分布的标度参数  $b_0$ 。

对于泊松先验，指定伽玛分布的标度参数  $b_0$ 。

必须输入一个大于 0 的值。

## 贝叶斯单样本推论：泊松

此功能需要 SPSS Statistics Standard Edition 或“高级统计”选项。

“贝叶斯单样本推论：泊松”过程提供用于对泊松分布执行贝叶斯单样本推论的选项。泊松分布（一种针对罕见事件的有用模型）假设在较小时间间隔内，事件发生的概率与等待时间的长度成比例。在得出对泊松分布的贝叶斯统计推论时，将使用伽玛分布族中的共轭先验。

1. 从菜单中选择：

分析 > 贝叶斯统计 > 泊松样本

2. 从可用变量列表中选择相应的检验变量。应至少选择一个变量。

注：可用变量列表提供除日期和字符串变量以外的所有变量。

3. 选择所需的贝叶斯分析：

- **描述后验分布特征：**选中该选项时，将通过描述后验分布的特征所得出的视角做出贝叶斯推论。您可以通过集成其他冗余参数来调查感兴趣参数的边际后验分布，然后进一步构造可信区间来得出直接推论。这是缺省设置。
- **估算贝叶斯因子：**如果选中该选项，那么在比较原假设和替代假设之间的边际似然估计时，估算贝叶斯因子（贝叶斯推论中的一种著名的方法）占据一定的自然比。

表 4: 常用于定义证据显著性的阈值

贝叶斯因子	证据类别	贝叶斯因子	证据类别	贝叶斯因子	证据类别
>100	H1 极端证据	1-3	H1 轶事证据	1/30-1/10	H0 有力证据
30-100	H1 超有力证据	1	无证据	1/100-1/30	H0 超有力证据
10-30	H1 有力证据	1/3-1	H0 轶事证据	1/100	H0 极端证据
3-10	H1 适度证据	1/10-1/3	H0 适度证据		

$H_0$ : 原假设

$H_1$ : 其他假设

5

6

- **使用两种方法：**如果选中该选项，那么将同时使用描述后验分布特征和估算贝叶斯因子推论方法。
4. 选择和/或输入相应的假设值设置。该表反映的是当前位于检验变量列表中的变量。随着检验变量中变量的添加或删除，该表会自动在其变量对列中添加或删除相同的变量。
- 当选择描述后验分布特征作为贝叶斯分析时，将不启用任何列。
  - 当选择估算贝叶斯因子或使用两种方法作为贝叶斯分析时，将启用所有可编辑列。

### 原点

启用和禁用原比率选项。启用此设置后，原先验形状和原先验刻度选项都将禁用。

### 原先验形状

在泊松推论的原假设下指定形状参数  $a_0$ 。

### 原先验刻度

在泊松推论的原假设下指定刻度参数  $b_0$ 。

### 原比率

在共轭先验分布的原假设下（以适应泊松-伽玛关系），指定形状参数  $a_0$  和刻度参数  $b_0$ 。最小值必须是大于 0 的数字值；最大值必须是最大的双精度值。

<sup>5</sup> Lee, M.D., and Wagenmakers, E.-J. 2013. *Bayesian Modeling for Cognitive Science: A Practical Course*. Cambridge University Press.

<sup>6</sup> Jeffreys, H. 1961. *Theory of probability*. Oxford University Press.

### 替代先验形状

必需参数，用于在泊松推论的备用假设下指定  $a_1$ （如果要估算贝叶斯因子）。

### 备用先验刻度

必需参数，用于在泊松推论的备用假设下指定  $b_1$ （如果要估算贝叶斯因子）。

- 您可以选择性地单击**标准**来指定第 83 页的『贝叶斯单一样本推论：标准』设置（可信区间百分比、缺失值选项和数值方法设置），或者单击**先验**来指定第 86 页的『贝叶斯单一样本推论：二项式/泊松先验』设置（共轭或定制先验分布）。

## 贝叶斯相关样本推论：正态

此功能需要 SPSS Statistics Standard Edition 或“高级统计”选项。

“贝叶斯相关样本推论：正态”过程为配对样本提供贝叶斯单样本推论选项。您可成对指定变量名称，并对平均值差值运行贝叶斯分析。

- 从菜单中选择：

分析 > 贝叶斯统计 > 相关样本正常

- 从**可用变量**列表中选择相应的**成对变量**。必须至少选择一对源变量，并且不能为任何给定配对集选择两个以上源变量。

**注：**可用变量列表提供除字符串变量以外的所有变量。

- 选择所需的**贝叶斯分析**：

- 描述后验分布特征：**选中该选项时，将通过描述后验分布的特征所得出的视角做出贝叶斯推论。您可以通过集成其他冗余参数来调查感兴趣参数的边际后验分布，然后进一步构造可信区间来得出直接推论。这是缺省设置。
- 估算贝叶斯因子：**如果选中该选项，那么在比较原假设和替代假设之间的边际似然估计时，估算贝叶斯因子（贝叶斯推论中的一种著名的方法）占据一定的自然比。

表 5: 常用于定义证据显著性的阈值

贝叶斯因子	证据类别	贝叶斯因子	证据类别	贝叶斯因子	证据类别
>100	H1 极端证据	1-3	H1 轶事证据	1/30-1/10	H0 有力证据
30-100	H1 超有力证据	1	无证据	1/100-1/30	H0 超有力证据
10-30	H1 有力证据	1/3-1	H0 轶事证据	1/100	H0 极端证据
3-10	H1 适度证据	1/10-1/3	H0 适度证据		

$H_0$ : 原假设

$H_1$ : 其他假设

7

8

- 使用两种方法：**如果选中该选项，那么将同时使用**描述后验分布特征**和**估算贝叶斯因子**推论方法。
- 选择和/或输入相应的**数据方差和假设值**设置。该表反映当前位于**成对变量**列表中的变量对。随着**配对变量**列表中变量对的添加或移除，该表自动在其变量对列中添加或移除相同的变量对。
  - 当**成对变量**列表中有一个或多个变量对时，将启用**方差已知**和**方差值**列。

### 方差已知

当方差为已知时，针对每个变量选择此选项。

<sup>7</sup> Lee, M.D., and Wagenmakers, E.-J. 2013. *Bayesian Modeling for Cognitive Science: A Practical Course*. Cambridge University Press.

<sup>8</sup> Jeffreys, H. 1961. *Theory of probability*. Oxford University Press.

### 方差值

可选参数，用于指定观察到的数据的方差值（如果已知）。

- 当**配对变量**列表中有一个或多个变量对，并且未选择**描述后验分布特征**时，将启用**空检验值**和**g 值**列。

### 零检验值

必需参数，用于在贝叶斯因子估算中指定空值。仅允许一个值，缺省值为 0。

### g 值

指定该值以在贝叶斯因子估算中定义  $\psi^2 = g\sigma^2_x$ 。指定**方差值**时，**g 值**缺省为 1。如果未指定**方差值**，那么可以指定固定 **g** 或省略将其集成的值。

5. 您可以选择性地单击**标准**来指定第 83 页的『贝叶斯单一样本推论：标准』设置（可信区间百分比、缺失值选项和数值方法设置），或者单击**先验**来指定第 86 页的『贝叶斯单一样本推论：二项式/泊松先验』设置（共轭或定制先验分布）。

## 贝叶斯独立样本推论

此功能需要 SPSS Statistics Standard Edition 或“高级统计”选项。

“贝叶斯独立样本推论”过程为使用组变量来定义两个不相关组以及对两个组平均值的差值执行贝叶斯推论提供选项。您可以使用不同方法估算贝叶斯因子，也可以通过假设方差为已知或未知来描述所需后验分布的特征。

1. 从菜单中选择：

分析 > 贝叶斯统计 > 独立样本正态分布

2. 从源变量列表中选择相应的**检验变量**。必须至少选择一个源变量。
3. 从**可用变量**列表中选择相应的**分组变量**。分组变量为未配对 *t* 检验定义两个组。所选分组变量可以是数字或字符串变量。
4. 选择所需的**贝叶斯分析**：
  - **描述后验分布特征**：选中该选项时，将通过描述后验分布的特征所得出的视角做出贝叶斯推论。您可以通过集成其他冗余参数来调查感兴趣参数的边际后验分布，然后进一步构造可信区间来得出直接推论。这是缺省设置。
  - **估算贝叶斯因子**：如果选中该选项，那么在比较原假设和替代假设之间的边际似然估计时，估算贝叶斯因子（贝叶斯推论中的一种著名的方法）占据一定的自然比。

表 6: 常用于定义证据显著性的阈值

贝叶斯因子	证据类别	贝叶斯因子	证据类别	贝叶斯因子	证据类别
>100	H1 极端证据	1-3	H1 轶事证据	1/30-1/10	H0 有力证据
30-100	H1 超有力证据	1	无证据	1/100-1/30	H0 超有力证据
10-30	H1 有力证据	1/3-1	H0 轶事证据	1/100	H0 极端证据
3-10	H1 适度证据	1/10-1/3	H0 适度证据		

*H0*: 原假设

*H1*: 其他假设

9

10

- **使用两种方法**：如果选中该选项，那么将同时使用**描述后验分布特征**和**估算贝叶斯因子**推论方法。
5. 使用**定义组**选项来定义两个用于 *t* 检验的组，方法是指定两个值（对于字符串变量），或者指定两个值、一个中点或一个分割点（对于数字变量）。

<sup>9</sup> Lee, M.D., and Wagenmakers, E.-J. 2013. *Bayesian Modeling for Cognitive Science: A Practical Course*. Cambridge University Press.

<sup>10</sup> Jeffreys, H. 1961. *Theory of probability*. Oxford University Press.

注: 指定的值必须存在于变量中, 否则会显示错误消息以指示至少其中一个组为空。

对于数字变量:

- **使用指定值。** 为组 1 输入一个值, 为组 2 输入另一个值。从分析中排除了所有其他值的个案。数字不需要是整数 (例如 6.25 和 12.5 也有效)。
- **使用中点值。** 选定时, 组分成  $<$  和  $\geq$  中点值。
- **使用分割点。**
  - **分割点。** 输入一个将分组变量的值分成两组的数字。值小于分割点的所有个案组成一个组, 值大于等于分割点的个案组成另一个组。

对于字符串分组变量, 为组 1 输入一个字符串, 为组 2 输入另一个值, 例如, *yes* 和 *no*。具有其他字符串的案被排除在分析之外。

6. 您可以选择性地单击**标准**来指定第 90 页的『贝叶斯独立样本推论: 标准』设置 (可信区间百分比、缺失值选项和自适应正交方法设置), 单击**先验**来指定第 91 页的『贝叶斯独立样本推论: 先验分布』设置 (数据方差、方差先验和视方差而定的平均值的先验), 或者单击**估算贝叶斯因子**以指定第 92 页的『贝叶斯独立样本推论: 估算贝叶斯因子』设置。

## 贝叶斯独立样本推论定义组 (数字)

对于数字分组变量, 请通过指定两个值 (中点或分割点) 来定义 *t* 测试的两个组。

注: 指定的值必须存在于变量中, 否则会显示错误消息以指示至少其中一个组为空。

- **使用指定值。** 为组 1 输入一个值, 为组 2 输入另一个值。从分析中排除了所有其他值的个案。数字不需要是整数 (例如 6.25 和 12.5 也有效)。
- **使用中点值。** 选定时, 组分成  $<$  和  $\geq$  中点值。
- **使用分割点。**
  - **分割点。** 输入一个将分组变量的值分成两组的数字。值小于分割点的所有个案组成一个组, 值大于等于分割点的个案组成另一个组。

## 贝叶斯独立样本推论定义组 (字符串)

对于字符串分组变量, 为组 1 输入一个字符串, 为组 2 输入另一个值, 例如, *yes* 和 *no*。具有其他字符串的案被排除在分析之外。

注: 指定的值必须存在于变量中, 否则会显示错误消息以指示至少其中一个组为空。

## 贝叶斯独立样本推论: 标准

您可以为贝叶斯独立样本推论指定以下分析标准:

### 可信区间百分比

指定用于计算可信区间的显著性水平。缺省级别为 95%。

### 缺失值

指定用于控制缺失值的方法。

### 成对排除案例

这是缺省设置, 用于按分析排除具有缺失值的记录。检验中将省略包含该特定检验所用字段的缺失值的记录。

### 成列排除案例

该设置按列表排除包含缺失值的记录。如果记录包含在任何子命令上指定的任何字段的缺失值, 那么会从所有分析中排除此类记录。

注: 仅当为贝叶斯分析选择**估算贝叶斯因子**或使用**两种方法**选项时, 以下选项才可用。

### 自适应正交方法

指定自适应正交方法的容差和最大迭代值。



### 容差

指定数值方法的容差值。缺省设置为 0.000001。

### 最大迭代数

指定最大自适应正交方法迭代次数。此值必须是正整数。缺省设置为 500。

## 贝叶斯独立样本推论：先验分布

您可以为贝叶斯独立样本推论指定以下先验分布标准：

**注：**很多应用研究员可能质疑指定先验的需求。参考先验可最大限度减少因为数据增加造成先验不堪重负的问题。在指定信息先验信息时，贝叶斯方法可有效地使用信息。指定先验的需要不应被视为妨碍使用贝叶斯分析。

### 数据方差

提供用于定义数据方差设置的选项。

#### 方差已知

选定 $\chi^2$ 时，允许输入两个已知组方差。这两个值必须  $> 0$ 。

#### 组 1 方差

输入第一个已知组方差值。

#### 组 2 方差

输入第二个已知组方差值。

#### 假设等方差

控制两个组方差是否假设为相等。缺省情况下，假设组方差不等。为两个组方差输入值之后，将忽略此设置。

#### 假设不等方差

控制两个组方差是否假设为不等。缺省情况下，假设组方差不等。为两个组方差输入值之后，将忽略此设置。

### 方差先验

指定两个等方差的先验分布。

#### Jeffreys

选定 $\chi^2$ 时，将使用参数空间的无信息（目标）先验分布。

#### 逆卡方

指定逆  $\chi^2(v_0, \sigma^2_0)$  的正值随机变量和参数的连续概率分布，其中  $v_0 > 0$  是自由度， $\sigma^2_0 > 0$  是刻度参数。

#### 自由度

在最终计算中指定自由变化的值的数量值。

#### 刻度参数

指定逆  $\chi^2(v_0, \sigma^2_0)$  的刻度参数  $\sigma^2_0 > 0$ 。必须输入一个大于 0 的值。刻度参数越大，分布越分散。

### 视方差而定的平均值的先验

提供多个用于指定两个组平均值的先验分布的选项。

**注：**仅当选择方差已知时，扩散和正态选项才可用。

#### 扩散

缺省设置。指定扩散先验。

#### 正态

选定 $\chi^2$ 时，必须指定已定义的组平均值的位置和刻度参数。

#### 位置参数

输入一个数字值，其指定组分布的位置参数。

#### 刻度参数

指定逆  $\chi^2(v_0, \sigma^2_0)$  的刻度参数  $\sigma^2_0 > 0$ 。对于每个组，必须输入大于 0 的单个值。刻度参数越大，分布越分散。

## 贝叶斯独立样本推论：估算贝叶斯因子

您可以指定用于估算贝叶斯因子的方法。

### Rouder 方法

选定时，调用 Rouder 方法。这是缺省设置

### Gonen 方法

选定时，调用 Gonen 方法，并且必须指定以下效应大小设置：

#### 效应大小的平均值

输入用于指定两个组之间的均值差的值。

#### 效应大小的方差

输入用于指定两个组的方差的值。该值必须  $> 0$ 。

### 超先验方法

选定时，调用需要指定单个值的 Hyper-g 方法。在形状参数字段中，输入介于 -1 与 -0.5 之间的值。缺省值为 -0.75。

## 关于 Pearson 相关性的贝叶斯推论

此功能需要 SPSS Statistics Standard Edition 或“高级统计”选项。

Pearson 相关系数测量联合遵循双变量正态分布的两个刻度变量之间的线性相关性。对于有关相关系数的传统统计推论已进行广泛讨论，并且在 IBM SPSS Statistics 中长期提供其实践。有关 Pearson 相关系数的贝叶斯推论的设计允许用户通过估算贝叶斯因子和描述后验分布特征来得出贝叶斯推论。

1. 从菜单中选择：

分析 > 贝叶斯统计 > 皮尔逊相关性

2. 从可用变量列表中选择相应的检验变量以用于成对相关推论。必须至少选择两个源变量。当选择两个以上变量时，将对所选变量的所有成对组合运行分析。

3. 选择所需的贝叶斯分析：

- **描述后验分布特征：**选中该选项时，将通过描述后验分布的特征所得出的视角做出贝叶斯推论。您可以通过集成其他冗余参数来调查感兴趣参数的边际后验分布，然后进一步构造可信区间来得出直接推论。这是缺省设置。
- **估算贝叶斯因子：**如果选中该选项，那么在比较原假设和替代假设之间的边际似然估计时，估算贝叶斯因子（贝叶斯推论中的一种著名的方法）占据一定的自然比。

贝叶斯因子	证据类别	贝叶斯因子	证据类别	贝叶斯因子	证据类别
$>100$	H1 极端证据	1-3	H1 轶事证据	1/30-1/10	H0 有力证据
30-100	H1 超有力证据	1	无证据	1/100-1/30	H0 超有力证据
10-30	H1 有力证据	1/3-1	H0 轶事证据	1/100	H0 极端证据
3-10	H1 适度证据	1/10-1/3	H0 适度证据		

$H_0$ : 原假设

$H_1$ : 其他假设

11

12

- **使用两种方法：**如果选中该选项，那么将同时使用描述后验分布特征和估算贝叶斯因子推论方法。

<sup>11</sup> Lee, M.D., and Wagenmakers, E.-J. 2013. *Bayesian Modeling for Cognitive Science: A Practical Course*. Cambridge University Press.

<sup>12</sup> Jeffreys, H. 1961. *Theory of probability*. Oxford University Press.

4. 指定要在输出中显示的**最大图数**。图集可在同一窗格上包含 3 个图。图按以下顺序生成：首先是第一个变量与剩余变量，然后是第二个变量与剩余变量，依此类推。所定义的整数值必须介于 0 和 50 之间。缺省情况下，将输出 10 个图集以容纳五个变量。选择**估算贝叶斯因子**时，此选项不可用。
5. 您可以选择单击**标准**以指定第 93 页的『贝叶斯 Pearson 相关性：条件』设置（可信时间间隔百分比，缺失值选项和数字方法设置），单击**Priors**以指定第 93 页的『贝叶斯 Pearson 相关性：先验分布』设置（对于先前的  $p(\rho) \propto (1 - \rho^2)^c$  指定值  $c$ ，或者单击**贝叶斯因子**以指定第 92 页的『贝叶斯独立样本推论：估算贝叶斯因子』设置。

## 贝叶斯 Pearson 相关性：条件

您可以为贝叶斯 Pearson 相关性推论指定以下分析条件（按对）。

### 可信区间百分比

指定用于计算可信区间的显著性水平。缺省级别为 95%。

### 缺失值

指定用于控制缺失值的方法。

#### 成对排除案例

此设置按对排除包含缺失值的记录。

#### 成列排除案例

该设置按列表排除包含缺失值的记录。如果记录包含在任何子命令上指定的任何字段的缺失值，那么会从所有分析中排除此类记录。

**注：**仅当为贝叶斯分析选择**估算贝叶斯因子**或使用**两种方法**选项时，以下选项才可用。

### 数值方法

指定用于估算积分的数值方法。

#### 设置定制种子

选中该选项时，可以在**种子值**字段中指定定制种子值。

#### 容差

指定数值方法的容差值。缺省设置为 0.000001。

#### 最大迭代数

指定最大方法迭代数。此值必须是正整数。缺省设置为 2000。

#### 蒙特卡洛样本的数量

指定为 Monte Carlo 近似值抽样的点数。此值必须是正整数。缺省值是 10000。

#### 针对后验分布模拟的样本数

指定用于抽取所需后验分布的样本数。缺省值是 10000。

## 贝叶斯 Pearson 相关性：先验分布

您可以为先验  $p(\rho) \propto (1 - \rho^2)^c$  指定值  $c$ 。

**注：**很多应用研究员可能质疑指定先验的需求。参考先验可最大限度减少因为数据增加造成先验不堪重负的问题。在指定信息先验信息时，贝叶斯方法可有效地使用信息。指定先验的需要不应被视为妨碍使用贝叶斯分析。

### 统一 ( $c = 0$ )

选定时，将使用统一先验。

### 杰弗里斯 ( $c = -1.5$ )

选定时，将使用无信息先验分布。

### 设置定制 $c$ 值

选定时，可以指定定制  $c$  值。允许任何单一实数。

## 贝叶斯 Pearson 相关性：贝叶斯因子

您可以指定用于估算贝叶斯因子的方法。仅当选择**估算贝叶斯因子**或使用**两种方法**贝叶斯分析选项时，以下选项才可用。

## JZS 贝叶斯因子

选定时，调用 Zellner-Siow 方法。这是缺省设置。

## 小数贝叶斯因子

选定时，您可以指定小数贝叶斯因子和原假设值。对于小数贝叶斯因子，必须指定值  $\in (0,1)$ 。缺省值为 0.5。

## 关于线性回归模型的贝叶斯推论

此功能需要 SPSS Statistics Standard Edition 或“高级统计”选项。

回归是定量建模中广泛使用的一种统计方法。线性回归是一种基本的标准方法，其中研究人员使用多个变量值来说明或预测刻度结果的值。贝叶斯单变量线性回归是在贝叶斯推论的上下文中执行统计分析的一种线性回归方法。

您可以调用回归过程并定义完整模型。

1. 从菜单中选择：

分析 > 贝叶斯统计 > 线性回归

2. 从可用变量列表中选择单个非字符串因变量。必须选择一个非字符串变量。

3. 从可用变量列表中为模型选择一个或多个分类因子变量。

4. 从可用变量列表选择一个或多个非字符串协变量刻度变量。

注：因子和协变量列表都不能为空。必须选择至少一个因子或协变量变量。

5. (可选) 从可用变量列表中选择要充当回归权重的单个非字符串变量。

6. 选择所需的贝叶斯分析：

- **描述后验分布特征：**选中该选项时，将通过描述后验分布的特征所得出的视角做出贝叶斯推论。您可以通过集成其他冗余参数来调查感兴趣参数的边际后验分布，然后进一步构造可信区间来得出直接推论。这是缺省设置。
- **估算贝叶斯因子：**如果选中该选项，那么在比较原假设和替代假设之间的边际似然估计时，估算贝叶斯因子（贝叶斯推论中的一种著名的方法）占据一定的自然比。

表 8: 常用于定义证据显著性的阈值

贝叶斯因子	证据类别	贝叶斯因子	证据类别	贝叶斯因子	证据类别
>100	H1 极端证据	1-3	H1 轶事证据	1/30-1/10	H0 有力证据
30-100	H1 超有力证据	1	无证据	1/100-1/30	H0 超有力证据
10-30	H1 有力证据	1/3-1	H0 轶事证据	1/100	H0 极端证据
3-10	H1 适度证据	1/10-1/3	H0 适度证据		

$H_0$ : 原假设

$H_1$ : 其他假设

13

14

- **使用两种方法：**如果选中该选项，那么将同时使用描述后验分布特征和估算贝叶斯因子推论方法。

(可选) 您可以执行以下操作：

- 单击**标准**以指定可信区间百分比和数值方法设置。
- 单击**先验**以定义参考和共轭先验分布设置。

<sup>13</sup> Lee, M.D., and Wagenmakers, E.-J. 2013. *Bayesian Modeling for Cognitive Science: A Practical Course*. Cambridge University Press.

<sup>14</sup> Jeffreys, H. 1961. *Theory of probability*. Oxford University Press.

- 单击**贝叶斯因子**以指定贝叶斯因子设置。
- 单击**保存**以标识要保存的项，并将模型信息保存到 XML 文件。
- 单击**预测**以指定贝叶斯预测的回归量。
- 单击**绘制**以绘制回归参数、误差项方差和预测值的后验分布。
- 单击**F 检验**来比较统计模型，以便识别与从中进行数据采样的群体最佳拟合的模型。

## 贝叶斯线性回归模型：条件

您可以为贝叶斯线性回归模型指定以下分析标准。

### 可信区间百分比

指定用于计算可信区间的显著性水平。缺省级别为 95%。

**注：**仅当为**贝叶斯分析**选择**估算贝叶斯因子**或**使用两种方法**选项时，以下选项才可用。

### 数值方法

指定用于估算积分的数值方法。

### 容差

指定数值方法的容差值。缺省设置为 0.000001。

### 最大迭代数

指定最大方法迭代数。此值必须是正整数。缺省设置为 2000。

## 贝叶斯线性回归模型：先验分布

您可以为回归参数和误差方差指定以下先验分布设置。仅当为**贝叶斯分析**选择**描述后验分布特征**选项时，以下选项才可用。

**注：**很多应用研究员可能质疑指定先验的需求。参考先验可最大限度减少因为数据增加造成先验不堪重负的问题。在指定信息先验信息时，贝叶斯方法可有效地使用信息。指定先验的需要不应被视为妨碍使用贝叶斯分析。

### 参考

选定时，参考分析生成目标贝叶斯推论。推论语句取决于采用的模型和可用数据，并且用于进行推论的先验分布信息最少。这是缺省设置。

### 共轭

提供用于定义共轭先验分布的选项。共轭先验采用正则逆伽玛联合分布。虽然在执行贝叶斯更新时无需共轭先验，但是它们有助于计算过程。

**注：**要为线性回归模型指定共轭先验，请在**误差方差的先验**表中设置期望的回归参数的平均值。您还可以选择使用**协方差矩阵变量**设置，以指定先验方差-协方差。

### 误差方差的先验

#### 形状参数

指定逆伽玛分布的形状参数  $a_0$ 。必须输入一个大于 0 的值。

#### 标度参数

指定逆伽玛分布的刻度参数  $b_0$ 。必须输入一个大于 0 的值。刻度参数越大，分布就越分散。

该表列出回归参数（包括截距）的平均值，该值指定已定义的回归参数的平均值向量  $\theta_0$ 。值的数量必须与回归参数的数量相符，包括截距项。

第一个变量名称始终是 INTERCEPT。从第二行起，变量列自动使用因子和协变量指定的变量进行填充。平均值列不包含任何缺省值。

单击**重置**可清除值。

### 协方差矩阵： $\sigma^2x$

在多变量正态先验的协方差矩阵中的下三角形中指定  $V_0$  值。请注意， $V_0$  必须为半正定。每行的最后一个值必须为正。下一行应比上一行多一个值。对于参考类别（如果有），不会指定任何值。

单击**重置**可清除值。

### 使用恒等矩阵

选定时，将使用已标度的恒等矩阵。不能在多变量正态先验的协方差矩阵中的下三角形中指定  $V_0$  值。

## 贝叶斯线性回归模型：贝叶斯因子

您可以为分析指定模型设计，包括用于估算贝叶斯线性回归模型的贝叶斯因子的方法。仅当选择**估算贝叶斯因子**或使用**两种方法**贝叶斯分析选项时，以下选项才可用。

### 空模型

选定时，估算的贝叶斯因子基于空模型。这是缺省设置。

### 完整模型

选定时，估算的贝叶斯因子基于完整模型，您可以选择要使用的变量以及附加因子和协变量。

#### 变量

列出可用于完整模型的所有变量。

#### 附加因子

从**变量**列表中选择要用作附加因子的变量。

#### 附加协变量

从**变量**列表中选择要用作附加协变量的变量。

### 计算

指定用于估算贝叶斯因子的方法。缺省设置为 JZS 方法。

#### JZS 方法

选定时，调用 Zellner-Siow 方法。这是缺省设置。

#### Zellner 方法

选定时，调用 Zellner 方法，并且您需要指定  $> 0$  的单个  $g$  先验值（没有缺省值）。

#### 超先验方法

选定时，调用 Hyper- $g$  方法，并且您需要为逆伽玛分布指定形状参数  $a_0$ 。必须指定  $> 0$  的单个值（缺省值为 3）。

#### Rouder 方法

选定时，调用 Rouder 方法，并且您需要为逆伽玛分布指定标度参数  $b_0$ 。必须指定  $> 0$  的单个值（缺省值为 1）。

## 贝叶斯线性回归模型：保存

通过此对话框，可以指定对贝叶斯预测分布的哪些统计信息进行评分，并将模型结果导出到 XML 文件。

### 后验预测性统计信息

可以对派生自贝叶斯预测的以下统计信息进行评分。

#### 均值

后验预测分布的均值。

#### 方差

后验预测分布的方差。

#### 方式

后验预测分布的方式。

#### 可信区间下限

后验预测分布的可信区间的下边界。

#### 可信区间上限

后验预测分布的可信区间的上边界。

**注：**您可以为每项统计信息分配对应的变量名称。

### 将模型信息导出到 XML 文件

输入 XML 文件名和位置来导出已评分的参数协方差矩阵。

## 贝叶斯线性回归模型：预测

您可以指定回归量来生成预测分布。

### 贝叶斯预测的回归量

该表列出所有可用回归量。回归量列通过因子和协变量变量自动填充。请使用回归量的值指定观察到的向量。可以为每个回归量分配一个值或字符串，并且允许其仅预测一种情况。对于因子，同时允许值和字符串。

必须指定所有回归量值或不得指定任何回归量值，以便运行预测（通过单击**运行分析**）。

移除因子或协变量后，将从表中移除对应的回归量行。

对于协变量，只能指定数字值。对于因子，同时允许数字值和字符串。

**注：**单击**重置**可清除已定义的值。

## 贝叶斯线性回归模型：图

您可以控制输入的图。

### 协变量

列出当前定义的协变量。

### 图协变量

从**协变量**列表中选择要绘制的协变量并将其添加到**绘制协变量**列表中。

### 因子

列出当前定义的因子。

### 绘制因子

从**因子**列表中选择要绘制的因子并将其添加到**绘制因子**列表中。

### 要绘制的最大类别数

选择要绘制的最大类别数（单个正整数）。该设置适用于所有因子。缺省情况下，针对每个因子绘制前 2 个级别。

### 包含图

#### 截距项

选定时，将绘制截距项。缺省情况下将取消选择该设置。

#### 误差的方差

选定时，将绘制误差的方差。缺省情况下将取消选择该设置。

#### 贝叶斯预测分布

选定时，将绘制预测分布。缺省情况下将取消选择该设置。仅当选择有效的回归量值时，才能选择该设置。

## 贝叶斯线性回归模型：F 检验

您可以创建一个或多个部分 F 检验。F 检验是检验统计在原假设下具有 F 分布的任意统计检验。F 检验常用于比较已拟合到数据集的统计模型，以便识别与从中进行数据采样的群体最佳拟合的模型。

### 可用变量

列出从“贝叶斯线性回归”主对话框中选择的因子和协变量变量。在从主对话框添加或删除因子和协变量时，将相应更新列表。

### 检验变量

从**可用变量**列表中选择要检验的因子/协变量变量并将其添加到**检验变量**列表中。

**注：**当未选择检验因子或协变量时，必须选择**包含截距项**选项。

### 检验变量和值

指定要检验的值。值的数量必须与原始模型中的参数数量匹配。指定值时，必须为截距项指定第一个值（假设在未显式定义时所有值都为 0）。

### 包含截距项

选定时，将在检验中包含截距项。缺省情况下，未选择该设置。

启用时，请使用**检验值**字段来指定值。

### 检验标签（可选）

您可以选择性地为每个检验指定标签。可以指定最大长度为 255 字节的字符串。仅允许每个 F 检验有一个标签。

## 贝叶斯单因素 ANOVA

此功能需要 SPSS Statistics Standard Edition 或“高级统计”选项。

“单因素 ANOVA”过程按照单因子变量（自变量）生成对定量因变量的单向方差检验分析。方差分析用于检验多个平均值相等的假设。SPSS Statistics 支持贝叶斯因子、共轭先验和非信息先验。

1. 从菜单中选择：

**分析 > 贝叶斯统计 > 单向 ANOVA**

2. 从**可用变量**列表中选择单个数值因变量。必须至少选择一个变量。

3. 从**可用变量**列表中为模型选择单个因子变量。必须至少选择一个因子变量。

4. 从**可用变量**列表中选择要充当回归**权重**的单个非字符串变量。**权重**变量字段可以为空。

5. 选择所需的**贝叶斯分析**：

- **描述后验分布特征**：选中该选项时，将通过描述后验分布的特征所得出的视角做出贝叶斯推论。您可以通过集成其他冗余参数来调查感兴趣参数的边际后验分布，然后进一步构造可信区间来得出直接推论。这是缺省设置。
- **估算贝叶斯因子**：如果选中该选项，那么在比较原假设和替代假设之间的边际似然估计时，估算贝叶斯因子（贝叶斯推论中的一种著名的方法）占据一定的自然比。

表 9: 常用于定义证据显著性的阈值

贝叶斯因子	证据类别	贝叶斯因子	证据类别	贝叶斯因子	证据类别
>100	H1 极端证据	1-3	H1 轶事证据	1/30-1/10	H0 有力证据
30-100	H1 超有力证据	1	无证据	1/100-1/30	H0 超有力证据
10-30	H1 有力证据	1/3-1	H0 轶事证据	1/100	H0 极端证据
3-10	H1 适度证据	1/10-1/3	H0 适度证据		

$H_0$ ：原假设

$H_1$ ：其他假设

15

16

- **使用两种方法**：如果选中该选项，那么将同时使用**描述后验分布特征**和**估算贝叶斯因子**推论方法。

（可选）您可以执行以下操作：

- 单击**标准**以指定可信区间百分比和数值方法设置。
- 单击**先验**以定义参考和共轭先验分布设置。
- 单击**贝叶斯因子**以指定贝叶斯因子设置。
- 单击**图**以控制输出的图。

### 贝叶斯单因素 ANOVA：标准

您可以为贝叶斯单因素 ANOVA 模型指定以下分析标准。

<sup>15</sup> Lee, M.D., and Wagenmakers, E.-J. 2013. *Bayesian Modeling for Cognitive Science: A Practical Course*. Cambridge University Press.

<sup>16</sup> Jeffreys, H. 1961. *Theory of probability*. Oxford University Press.



## 可信区间百分比

指定用于计算可信区间的显著性水平。缺省级别为 95%。

注: 仅当为贝叶斯分析选择估算贝叶斯因子或使用两种方法选项时, 以下选项才可用。

## 数值方法

指定用于估算积分的数值方法。

### 容差

指定数值方法的容差值。缺省设置为 0.000001。

### 最大迭代数

指定最大方法迭代数。此值必须是正整数。缺省设置为 2000。

## 贝叶斯单因素 ANOVA: 先验

您可以为回归参数和误差方差指定以下先验分布设置。仅当为贝叶斯分析选择描述后验分布特征选项时, 以下选项才可用。

注: 很多应用研究员可能质疑指定先验的需求。参考先验可最大限度减少因为数据增加造成先验不堪重负的问题。在指定信息先验信息时, 贝叶斯方法可有效地使用信息。指定先验的需要不应被视为妨碍使用贝叶斯分析。

### 参考

选定时, 参考分析生成目标贝叶斯推论。推论语句取决于采用的模型和可用数据, 并且用于进行推论的先验分布信息最少。这是缺省设置。

### 共轭

提供用于定义共轭先验分布的选项。共轭先验采用正则逆伽玛联合分布。虽然在执行贝叶斯更新时无需共轭先验, 但是它们有助于计算过程。

### 误差方差的先验

#### 形状参数

指定逆伽玛分布的形状参数  $\alpha_0$ 。必须输入一个大于 0 的值。

#### 标度参数

指定逆伽玛分布的刻度参数  $b_0$ 。必须输入一个大于 0 的值。刻度参数越大, 分布越分散。

### 先验选择回归参数

指定组平均值的均值向量  $\beta_0$ 。值的数量必须与回归参数的数量相符, 包括截距项。

变量列自动使用因子的级别进行填充。平均值列不包含任何缺省值。

单击重置可清除值。

### 协方差矩阵: $\sigma^2\mathbf{x}$

在多变量正态先验的协方差矩阵中的下三角形中指定  $V_0$  值。请注意,  $V_0$  必须为半正定。必须仅指定表的下三角形。

行和列自动使用因子的级别进行填充。所有对角线值都为 1; 所有非对角线值都为 0。

单击重置可清除值。

### 使用恒等矩阵

选定时, 将使用恒等矩阵。不能在多变量正态先验的协方差矩阵中的下三角形中指定  $V_0$  值。

## 贝叶斯单因素 ANOVA: 贝叶斯因子

您可以指定用于估算贝叶斯单因素 ANOVA 模型的贝叶斯因子的方法。仅当选择估算贝叶斯因子或使用两种方法贝叶斯分析选项时, 以下选项才可用。

### 计算

指定用于估算贝叶斯因子的方法。缺省设置为 JZS 方法。

#### JZS 方法

选定时, 调用 Zellner-Siow 方法。这是缺省设置。

### Zellner 方法

选定时，调用 Zellner 方法，并且您需要指定  $> 0$  的单个  $g$  先验值（没有缺省值）。

### 超先验方法

选定时，调用 Hyper- $g$  方法，并且您需要为逆伽玛分布指定形状参数  $a_0$ 。必须指定  $> 0$  的单个值（缺省值为 3）。

### Rouder 方法

选定时，调用 Rouder 方法，并且您需要为逆伽玛分布指定标度参数  $b_0$ 。必须指定  $> 0$  的单个值（缺省值为 1）。

## 贝叶斯单因素 ANOVA：图

您可以控制输入的图。

### 图组

指定要绘制的子组。绘制指定组的平均值的似然、先验和后验。**组列表**是因子变量的类别的子集，因此格式应与因子的数据类型和实际值一致。

### 误差项的方差

选定时，将绘制误差的方差。缺省情况下将取消选择该设置。选择**估算贝叶斯因子**作为贝叶斯分析时，此选项不可用。

## 贝叶斯对数线性模型

此功能需要 SPSS Statistics Standard Edition 或“高级统计”选项。

用于检验两个因子的独立性的设计需要两个分类变量来构造列联表，以及对行-列关联进行贝叶斯推论。您可以通过采用不同模型来估算贝叶斯因子，并通过模拟交互项的同时可信区间来描述所需后验分布的特征。

1. 从菜单中选择：

分析 > 贝叶斯统计 > 对数线性模型

2. 从**可用变量**列表中选择单个非标度行变量。必须至少选择一个非标度变量。

3. 从**可用变量**列表中选择单个、非刻度列变量。必须至少选择一个非标度变量。

4. 选择所需的**贝叶斯分析**：

- **描述后验分布特征**：选中该选项时，将通过描述后验分布的特征所得出的视角做出贝叶斯推论。您可以通过集成其他冗余参数来调查感兴趣参数的边际后验分布，然后进一步构造可信区间来得出直接推论。这是缺省设置。
- **估算贝叶斯因子**：如果选中该选项，那么在比较原假设和替代假设之间的边际似然估计时，估算贝叶斯因子（贝叶斯推论中的一种著名的方法）占据一定的自然比。

表 10: 常用于定义证据显著性的阈值

贝叶斯因子	证据类别	贝叶斯因子	证据类别	贝叶斯因子	证据类别
>100	H1 极端证据	1-3	H1 轶事证据	1/30-1/10	H0 有力证据
30-100	H1 超有力证据	1	无证据	1/100-1/30	H0 超有力证据
10-30	H1 有力证据	1/3-1	H0 轶事证据	1/100	H0 极端证据
3-10	H1 适度证据	1/10-1/3	H0 适度证据		

$H_0$ ：原假设

$H_1$ ：其他假设

17

18

<sup>17</sup> Lee, M.D., and Wagenmakers, E.-J. 2013. *Bayesian Modeling for Cognitive Science: A Practical Course*. Cambridge University Press.

- **使用两种方法**：如果选中该选项，那么将同时使用**描述后验分布特征**和**估算贝叶斯因子推论方法**。

(可选) 您可以执行以下操作：

- 单击**标准**以指定可信区间百分比和数值方法设置。
- 单击**贝叶斯因子**以指定贝叶斯因子设置。
- 单击**打印**以指定内容在输出表中的显示方式。

## 贝叶斯对数线性模型：标准

您可以为贝叶斯对数线性模型指定以下分析标准。

### 可信区间百分比

指定用于计算可信区间的显著性水平。缺省级别为 95%。

### 数值方法

指定用于估算积分的数值方法。

### 设置定制种子

选中该选项时，可以在**种子值**字段中指定定制种子值。指定随机种子集合值。此值必须是正整数。缺省情况下，分配随机种子值。

**注**：仅当为**贝叶斯分析**选择**估算贝叶斯因子**或**使用两种方法**选项时，以下选项才可用。

### 容差

指定数值方法的容差值。缺省设置为 0.000001。

### 最大迭代数

指定最大方法迭代数。此值必须是正整数。缺省设置为 2000。

### 模拟为后验分布的样本数

指定用于抽取所需后验分布的样本数。缺省值是 10000。

### Format

选择按**升序**还是**降序**显示类别。缺省设置为“升序”。

## 贝叶斯对数线性模型：贝叶斯因子

您可以指定针对观察到的数据假设的模型（泊松、多项或非参数）。多项分布是缺省设置。仅当选择**估算贝叶斯因子**或**使用两种方法**贝叶斯分析选项时，以下选项才可用。

### 泊松模型

选定时，针对观察到的数据假设泊松模型。

### 多项模型

选定时，针对观察到的数据假设多项模型。这是缺省设置。

### 固定边际

选择**总计**、**行总和**或**列总和**以指定列联表的固定边际总计。**总计**是缺省设置。

### 先验分布

在估算贝叶斯因子时指定先验分布类型。

### 共轭

选择以指定共轭先验分布。使用**形状参数表**可指定伽玛分布的形状参数  $a_{rs}$ 。选择**共轭**作为先验分布类型时，必须指定形状参数。

当指定单个值时，假定所有  $a_{rs}$  都等于此值。 $a_{rs} = 1$  是缺省设置。如果需要指定多个值，那么可以使用空格分隔各个值。

每行和每列中指定的数字值的数量必须与列联表的维度匹配。所有指定的值都必须  $> 0$ 。

单击**重置**可清除值。

### 标度参数

指定伽玛分布的刻度参数  $b$ 。必须指定  $> 0$  的单个值。

<sup>18</sup> Jeffreys, H. 1961. *Theory of probability*. Oxford University Press.

### 混合 Dirichlet

选择以指定混合 Dirichlet 先验分布。

### 内在

选择以指定固有先验分布。

### 非参数模型

选定时，针对观察到的数据假设非参数模型。

### 固定边际

选择行总和或列总和以指定列联表的固定边际总计。行总和是缺省设置。

### 先验分布

指定 Dirichlet 先验的参数。选择非参数模型时，必须指定先验分布参数。当指定单个值时，假定所有  $\lambda_s$  等于该值。 $\lambda_s = 1$  是缺省设置。如果需要指定多个值，那么可以使用空格分隔各个值。所有指定的值都必须  $> 0$ 。所指定数字值的数量必须与针对列联表并非固定的行或列的维度匹配。

单击重置可清除值。

## 贝叶斯对数线性模型：打印

您可以指定内容在输出表中的显示方式。

### 表设计

#### 隐藏表

选定时，将不在输出中包含列联表。缺省情况下未启用此设置。

注：启用隐藏表设置后，以下设置无效。

### 统计量

指定用于检验独立性的统计信息。

#### 卡方

选择以计算 Pearson 卡方统计量、自由度和双尾渐进显著性。对于  $2 \times 2$  列联表，此设置还计算连续性校正统计量、自由度和关联的双尾渐进显著性。对于至少一个预期单元格计数  $< 5$  的  $2 \times 2$  列联表，此设置还计算 Fisher 的精确检验的双尾和单尾精确显著性。

#### 似然比

选择以计算似然比检验统计量、自由度和关联的双尾渐进显著性。

### 计数

指定在列联表中包含的计数类型。

#### 观测值

选择以在列联表中包含观察到的单元格计数。

#### 期望值

选择以在列联表中包含预期单元格计数。

### 百分比

指定在列联表中包含的百分比类型。

#### 行

选择以在列联表中包含行百分比。

#### 列

选择以在列联表中包含列百分比。

#### 合计

选择以在列联表中包含总百分比。

## 贝叶斯单向重复测量 ANOVA 模型

此功能需要 SPSS Statistics Standard Edition 或“高级统计”选项。

在贝叶斯单向方差检验分析 (ANOVA) 模型中，假设每个主体存在单个测量。但此假设并非始终正确。旨在研究平均值的调研对多个时间点或多种条件出现响应的情况并不少见。贝叶斯单向重复测量 ANOVA 过程会

在每个独立的时间点或条件中测量来自相同主体的一个因子，并允许这些主体在多个级别内交叉。假定每个主体针对每个时间点或条件都进行单次观察（因此，不考虑主体处理交互）。

1. 从菜单中选择：

分析 > 贝叶斯统计 > 单向重复测量 ANOVA

2. 从可用变量列表中选择至少两个重复测量变量。

3. (可选) 从可用变量列表中选择单个变量以充当回归权重。权重变量字段可以为空。

注：可用变量列表提供除字符串变量以外的所有变量。

4. 选择所需的贝叶斯分析：

- **描述后验分布特征**：选中该选项时，将通过描述后验分布的特征所得出的视角做出贝叶斯推论。您可以通过集成其他冗余参数来调查感兴趣参数的边际后验分布，然后进一步构造可信区间来得出直接推论。这是缺省设置。
- **估算贝叶斯因子**：如果选中该选项，那么在比较原假设和替代假设之间的边际似然估计时，估算贝叶斯因子（贝叶斯推论中的一种著名的方法）占据一定的自然比。

贝叶斯因子	证据类别	贝叶斯因子	证据类别	贝叶斯因子	证据类别
>100	H1 极端证据	1-3	H1 轶事证据	1/30-1/10	H0 有力证据
30-100	H1 超有力证据	1	无证据	1/100-1/30	H0 超有力证据
10-30	H1 有力证据	1/3-1	H0 轶事证据	1/100	H0 极端证据
3-10	H1 适度证据	1/10-1/3	H0 适度证据		

$H_0$ ：原假设

$H_1$ ：其他假设

19

20

- **使用两种方法**：如果选中该选项，那么将同时使用描述后验分布特征和估算贝叶斯因子推论方法。

(可选) 您可以执行以下操作：

- 单击**标准**以指定可信区间百分比和数值方法设置。
- 单击**贝叶斯因子**以指定贝叶斯因子设置。
- 单击**绘制**以绘制组均值的后验分布。

## 贝叶斯单因子重复测量变异数分析：条件

您可以为贝叶斯单向重复测量 ANOVA 模型指定以下分析标准。

### 可信区间百分比

指定用于计算可信区间的显著性水平。缺省级别为 95%。

### 数值方法

指定用于估算积分的数值方法。

### 设置定制种子

选中该选项时，可以在**种子值**字段中指定定制种子值。缺省值为 2,000,000。此值必须是介于 1 与 2,147,483,647 之间的正整数。缺省情况下，分配随机种子值。

<sup>19</sup> Lee, M.D., and Wagenmakers, E.-J. 2013. *Bayesian Modeling for Cognitive Science: A Practical Course*. Cambridge University Press.

<sup>20</sup> Jeffreys, H. 1961. *Theory of probability*. Oxford University Press.

### 蒙特卡洛样本的数量

指定为 Monte Carlo 近似值抽样的点数。此值必须是介于  $10^3$  与  $10^6$  之间的正整数。缺省值为 30,000。

## 贝叶斯单向重复测量 ANOVA：贝叶斯因子

您可以指定用于估算贝叶斯单向重复度量 ANOVA 模型的贝叶斯因子的方法。仅当选择**估算贝叶斯因子或使用两种方法**贝叶斯分析选项时，以下选项才可用。

### 贝叶斯信息标准 (BIC)

对重复测量设计使用 BIC 近似值的扩展来估算贝叶斯因子。此设置可派生出对应于重复测量关联的有效样本大小，并在为两个竞争模型之间的选择估算 BIC 时提供改进的惩罚项建议。这是缺省设置。

### Rouder 的混合设计

使用 Cauchy 分布的多变量泛化关系作为标准化效应大小的先验，并针对方差使用无信息先验。

注：选中该选项时，将忽略全局频率权重设置和回归权重。

## 贝叶斯单向重复测量 ANOVA：绘图

您可控制输出后显示组均值的后验分布的各种绘图。该表列示“变量”对话框上选中作为重复测量的所有变量。选择要绘制的重复测量变量。

## 内核岭回归

内核岭回归是一个扩展过程，它使用 Python `sklearn.kernel_ridge.KernelRidge` 类来估计内核岭回归模型。内核岭回归模型是一种非参数回归模型，能够对预测变量与结果之间的线性关系和非线性关系进行建模。结果可能对模型超参数的选择高度敏感。内核岭回归通过使用 `sklearn.model_selection.GridSearchCV` 类在所指定的值网格上进行 k 折交叉验证来促进超参数值的选择。

### 示例

#### 统计量

Additive\_CHI2、CHI2、余弦、拉普拉斯算子、线性、多项式、RBF、Sigmoid、Alpha、伽玛、Coef0、度数、交叉验证、实测与预测、残差与预测、双权重系数和内核空间权重系数。

## 数据注意事项

### 数据

- 您可以指定八个不同的内核函数中的任何一个或全部。
- 所选择的内核函数确定哪些超参数处于活动状态。
- 超参数包括所有内核通用的岭正则化 `alpha` 以及每个特定内核函数的多达三个其他超参数。
- 如果指定了多个内核子命令，或者指定了任何参数的多个值，那么将执行具有交叉验证的网格搜索以评估模型，并选择基于保留数据的最佳拟合模型。
- 该扩展接受来自“分割文件”过程的分割变量和使用“加权个案”过程的权重。
- 当包含权重时，它们用于在所有分析中创建拟合值。由于 `sklearn.model_selection.GridSearchCV` 类中的 `score` 方法的限制，用于模型选择的交叉验证评估没有加权。

### 假定

## 获得内核岭回归

1. 从菜单中选择：  
    分析 > 回归 > 核心岭...
2. 选择相关变量。

3. 选择一个或多个**独立**变量。
4. 当每个内核函数参数仅指定一个值时，将使用缺省的**单一模型**设置。选择了**单一模型**设置时，您无法指定其他**内核函数**，并且在结果的整个分析、评估和评分过程中将充分应用权重。您还可以使用向上箭头和向下箭头控件来重新排列内核函数。

(可选) 从**模式**列表中选择**模型选择**。

从**模式**列表中选择了**模型选择**时，可以将多个内核函数添加到**内核**列表。

- a. 单击添加控件 (+) 以包含其他内核函数。
- b. 单击“**内核**”列中的空单元格以选择内核函数。
- c. 双击任何核函数行单元格以指定相应列的核函数参数值 (**Alpha**、**Gamma**、**Coef0** 和 **Degree**)。有关更多信息，请参阅第 105 页的『**内核参数**』。下面列出了缺省内核函数调整参数。

#### **Additive\_CHI2**

ALPHA=1 GAMMA=1

#### **CHI2**

ALPHA=1 GAMMA=1

#### **余弦**

ALPHA=1

#### **拉普拉斯算子**

ALPHA=1 GAMMA=1/p

#### **线性**

缺省内核函数。ALPHA=1

#### **多项式**

ALPHA=1 GAMMA=1/p COEF0=1 DEGREE=3

#### **RBF**

ALPHA=1 GAMMA=1/p

#### **Sigmoid**

ALPHA=1 GAMMA=1/p COEF0=1

**注:** 如果为任何内核函数参数指定了多个值，那么将执行具有交叉验证的网格搜索以评估模型，并选择基于保留数据的最佳拟合模型。

5. (可选) 单击“**选项**”以指定交叉验证折叠数、显示选项、绘图设置和要保存的项。有关更多信息，请参阅第 105 页的『**核岭回归：选项**』。
6. 单击**确定**。

## **内核参数**

“**内核参数**”对话框提供了用于指定单个内核函数参数值的选项，以及用于指定使用对内核和指定网格参数值的组合进行网格搜索来执行模型选择的选项。

### **指定单个参数(F)**

启用此设置可指定所选内核函数参数的值。

- 输入一个值，然后单击“**添加**”以在内核函数参数中包含该值。
- 选择参数值，然后单击“**更改**”以更新该值。
- 选择参数值，然后单击“**移除**”以删除该值。

### **指定网格参数(P)**

启用此设置以指定使用对内核和指定网格参数值的组合进行网格搜索来执行模型选择。

## **核岭回归：选项**

“**绘图**”对话框提供用于指定交叉验证折数、显示选项、绘图设置和要保存的项的选项。

### 交叉验证折数

通过网格搜索进行交叉验证以进行模型选择的分割数或折数。输入大于 1 的整数值。默认值为 5。只有在主“模型选择”选择为模式时，此设置才可用。“核心岭回归”对话框上将

### 显示

提供用于指定在交叉验证生效时要显示的输出的选项。

#### 最佳

缺省设置仅显示所选最佳模型的基本结果。

#### 比较

显示所有已评估模型的基本结果。

#### 比较模型和折(F)

为每个评估模型显示每个分割或折叠的完整结果。

### 统计图

提供用于指定观测值或残差值与预测值的图的选项。

#### 实测值与预测值(O)

显示所指定模型或最佳模型的观测值与预测值的散点图。

#### 残差与预测值(R)

显示残差的散点图与所指定模型或最佳模型的预测值。

### 保存

该表提供了用于指定要保存到活动数据集中的变量的选项。

#### 预测值

将所指定模型或最佳模型中的预测值保存到活动数据集。可以包含可选变量名。

#### 残差

将所指定模型或最佳模型预测的残差保存到活动数据集中。可以包含可选变量名。

#### 双系数

将双或内核空间权重系数从指定模型保存到活动数据集。可以包含可选变量名。当在主“模型选择”选择为模式时，此设置不可用。“核心岭回归”对话框上将



# 声明

---

本信息是为在美国提供的产品和服务编写的。IBM 提供了本资料的其他语言版本。但是，您可能需要拥有该语言的产品副本或产品版本才能访问这些资料。

IBM 可能在其他国家或地区不提供本文档中所讨论的产品、服务或功能特性。有关您当前所在区域的产品和服务的信息，请向您当地的 IBM 代表咨询。任何对 IBM 产品、程序或服务的引用并非意在明示或暗示只能使用 IBM 的产品、程序或服务。只要不侵犯 IBM 的知识产权，任何同等功能的产品、程序或服务都可以代替 IBM 产品、程序或服务。但是，评估和验证任何非 IBM 产品、程序或服务，则由用户自行负责。

IBM 可能已拥有或正在申请与本文档内容有关的各项专利。提供本文档并未授予用户使用这些专利的任何许可。您可以通过书面方式将许可查询寄往：

*IBM Director of Licensing*

*IBM Corporation*

*North Castle Drive, MD-NC119  
Armonk, NY 10504-1785  
US*

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

*Intellectual Property Licensing  
Legal and Intellectual Property Law  
IBM Japan Ltd.*

*19-21, Nihonbashi-Hakozakicho, Chuo-ku  
Tokyo 103-8510, Japan*

International Business Machines Corporation“按现状”提供本出版物，不附有任何种类的（无论是明示的还是默示的）保证，包括但不限于默示的有关非侵权、适销和适用于某种特定用途的保证。某些管辖区域在某些交易中不允许免除明示或默示的保证。因此本条款可能不适用于您。

本信息可能包含技术方面不够准确的地方或印刷错误。此处的信息会定期进行更改；这些更改会体现在本出版物的新版本中。IBM 可以随时对本出版物中描述的产品进行改进和/或更改，而不另行通知。

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

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

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

*IBM Director of Licensing*

*IBM Corporation*

*North Castle Drive, MD-NC119  
Armonk, NY 10504-1785  
US*

只要遵守适当的条款和条件，包括某些情形下的一定数量的付费，都可获得这方面的信息。

本文档中描述的许可程序及其所有可用的许可资料均由 IBM 依据 IBM 客户协议、IBM 国际程序许可协议或任何同等协议中的条款提供。

所引用的性能数据和客户示例只用于阐述说明。根据具体配置和操作条件，实际性能结果可能有所不同。

涉及非 IBM 产品的信息是从这些产品的供应商、已出版说明或其他可公开获得的资料中获取。IBM 没有对这些产品进行测试，也无法确认其性能的精确性、兼容性或任何其他关于非 IBM 产品的声明。有关非 IBM 产品性能的问题应当向这些产品的供应商提出。

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

本信息包含日常业务运营中使用的数据和报告的示例。为了尽可能详尽地对其进行说明，示例中包含了人员的姓名、公司、品牌和产品的名称。所有这些名称都是虚构的，若实际人员或企业与此相似，纯属巧合。

版权许可：

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

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

© Copyright IBM Corp. 2021. 此部分代码是根据 IBM Corp. 的样本程序衍生出来的。

© Copyright IBM Corp. 1989 - 2021. All rights reserved.

## 商标

---

IBM、IBM 徽标和 [ibm.com](http://ibm.com) 是 International Business Machines Corp. 在全球许多司法辖区的商标或注册商标。其他产品和服务名称可能是 IBM 或其他公司的商标。最新的 IBM 商标列表可以在 Web 上的“Copyright and trademark information”中获取，地址为：[www.ibm.com/legal/copytrade.shtml](http://www.ibm.com/legal/copytrade.shtml)。

Adobe、Adobe 徽标、PostScript 和 PostScript 徽标是 Adobe Systems Incorporated 在美国和/或其他国家或地区的注册商标或商标。

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

Linux 是 Linus Torvalds 在美国和/或其他国家或地区的注册商标。

Microsoft、Windows、Windows NT 和 Windows 徽标是 Microsoft Corporation 在美国和/或其他国家或地区的商标。

UNIX 是 The Open Group 在美国和/或其他国家或地区的注册商标。

Java 和所有基于 Java 的商标和徽标是 Oracle 和/或其子公司的商标或注册商标。

# 索引

## Special Characters

### 饱和模型

在“模型选择对数线性分析”中 [48](#)

### 比例风险模型

在“Cox 回归”中 [71](#)

### 标准化残差

在“GLM 重复测量”中 [12](#)

在“GLM”中 [5](#)

### 标准误差

在“GLM 重复测量”中 [12](#)

在“GLM”中 [5](#)

### 参考类别

在“广义估算方程”中 [32](#)

在广义估算方程中 [33](#)

在广义线性模型中 [24](#)

### 参数 FRailty 模型

生存状态变量 [65](#)

### 参数估计

在“广义估计方程”中 [35](#)

在“模型选择对数线性分析”中 [49](#)

在“线性混合模型”中 [20](#)

在“一般对数线性分析”中 [49](#)

在“Logit 对数线性分析”中 [51](#)

在广义线性模型中 [26](#)

### 参数化共享脆弱模型

标准 [61](#)

打印 [64](#)

导出 [65](#)

分析 [60](#)

估算 [62](#)

绘制 [65](#)

模型 [62](#)

预测 [64](#)

### 参数加速故障时间模型

标准 [56](#)

打印 [58](#)

导出 [59](#)

分析 [55](#)

估算 [57](#)

绘制 [59](#)

模型 [56](#)

预测 [58](#)

### 参数收敛

在“广义估算方程”中 [33](#)

在“线性混合模型”中 [20](#)

在广义线性模型中 [25](#)

### 参数协方差矩阵

在“线性混合模型”中 [20](#)

### 残差

在“广义估计方程”中 [36](#)

在“模型选择对数线性分析”中 [49](#)

在“线性混合模型”中 [21](#)

在“一般对数线性分析”中 [51](#)

在“Logit 对数线性分析”中 [53](#)

在广义线性模型中 [27](#)

### 残差协方差矩阵

### 残差协方差矩阵 (继续)

在“线性混合模型”中 [20](#)

### 迭代

在“广义估算方程”中 [33](#)

在“模型选择对数线性分析”中 [49](#)

在广义线性模型中 [25](#)

### 迭代历史记录

在“广义估计方程”中 [35](#)

在“线性混合模型”中 [20](#)

在广义线性模型中 [26](#)

### 定制模型

在“方差成分”中 [14](#)

在“模型选择对数线性分析”中 [48](#)

在“GLM 重复测量”中 [9](#)

### 对比

在“一般对数线性分析”中 [49](#)

在“Cox 回归”中 [72](#)

在“Logit 对数线性分析”中 [51](#)

### 对比系数矩阵

在“广义估计方程”中 [35](#)

在广义线性模型中 [26](#)

### 对数等级检验

在“Kaplan-Meier”中 [70](#)

### 对数关联函数

在广义估计方程中 [30](#)

在广义线性模型中 [22](#)

### 对数互补关联函数

在广义估计方程中 [30](#)

在广义线性模型中 [22](#)

### 对数似然估计收敛性

在“广义估算方程”中 [33](#)

在“线性混合模型”中 [20](#)

在广义线性模型中 [25](#)

### 对数线性分析

分对数的对数线性分析 [51](#)

广义线性混合模型 [37](#)

一般对数线性分析 [49](#)

### 多变量 ANOVA [1](#)

### 多变量 GLM [1](#)

### 多变量回归 [1](#)

### 多变量模型

广义线性混合模型 [37](#)

### 多项 Logistic 回归

广义线性混合模型 [37](#)

### 多项分布

在广义估计方程中 [30](#)

在广义线性模型中 [22](#)

### 多项式 Logit 模型 [51](#)

### 二项分布

在广义估计方程中 [30](#)

在广义线性模型中 [22](#)

### 方差成分

保存结果 [15](#)

命令附加功能 [16](#)

模型 [14](#)

options [15](#)

### 方差分析

- 方差分析 (继续)
  - 广义线性混合模型 [37](#)
  - 在“方差成分”中 [15](#)
- 分层对数线性模型 [47](#)
- 分层解构
  - 在“方差成分”中 [15](#)
- 分层模型
  - 广义线性混合模型 [37](#)
- 分段依时协变量
  - 在“Cox 回归”中 [74](#)
- 分对数的对数线性分析
  - 保存变量 [53](#)
  - 标准 [53](#)
  - 残差 [53](#)
  - 单元格计数分布 [51](#)
  - 单元格结构 [51](#)
  - 单元格协变量 [51](#)
  - 对比 [51](#)
  - 绘图 [53](#)
  - 显示选项 [53](#)
  - 因子 [51](#)
  - 预测值 [53](#)
  - 置信区间 [53](#)
- 分离
  - 在“广义估算方程”中 [33](#)
  - 在广义线性模型中 [25](#)
- 风险率
  - 在“寿命表”中 [53](#)
- 负二项式分布
  - 在广义估计方程中 [30](#)
  - 在广义线性模型中 [22](#)
- 负二项式关联函数
  - 在广义估计方程中 [30](#)
  - 在广义线性模型中 [22](#)
- 负重对数关联函数
  - 在广义估计方程中 [30](#)
  - 在广义线性模型中 [22](#)
- 概要图
  - 在“GLM 多变量”中 [3](#)
  - 在“GLM 重复测量”中 [10](#)
- 杠杆值
  - 在“GLM 重复测量”中 [12](#)
  - 在“GLM”中 [5](#)
  - 在广义线性模型中 [27](#)
- 个案处理摘要
  - 在“广义估计方程”中 [35](#)
  - 在广义线性模型中 [26](#)
- 构建项 [2](#), [9](#), [15](#), [48](#), [50](#), [53](#)
- 估算边际平均值
  - 在“线性混合模型”中 [21](#)
  - 在广义估计方程中 [35](#)
  - 在广义线性模型中 [27](#)
- 固定效应
  - 在“线性混合模型”中 [18](#)
- 固定预测值
  - 在“线性混合模型”中 [21](#)
- 关联函数
  - 广义线性混合模型 [39](#)
- 广义对数几率比
  - 在“一般对数线性分析”中 [49](#)
- 广义估算方程
  - 初始值 [34](#)
  - 二元响应的参考类别 [32](#)
  - 分类因子的选项 [33](#)
- 广义估算方程 (继续)
  - 估计条件 [33](#)
  - 估算边际平均值 [35](#)
  - 将变量保存到活动数据集 [36](#)
  - 模型导出 [36](#)
  - 模型规范 [33](#)
  - 模型类型 [30](#)
  - 统计量 [35](#)
  - 响应 [32](#)
  - 预测变量 [32](#)
- 广义可估函数
  - 在“广义估计方程”中 [35](#)
- 广义线性混合模型
  - 保存字段 [44](#)
  - 定制项 [41](#)
  - 分类表 [45](#)
  - 分析权重 [42](#)
  - 估计均值 [47](#)
  - 估算边际平均值 [43](#)
  - 固定系数 [46](#)
  - 固定效应 [40](#), [45](#)
  - 关联函数 [39](#)
  - 模型导出 [44](#)
  - 模型视图 [45](#)
  - 模型摘要 [45](#)
  - 目标分布 [39](#)
  - 偏移 [42](#)
  - 数据结构 [45](#)
  - 随机效应 [41](#)
  - 随机效应块 [41](#)
  - 随机效应协方差 [46](#)
  - 协方差参数 [46](#)
  - 预测-实测 [45](#)
- 广义线性模型
  - 初始值 [26](#)
  - 二元响应的参考类别 [24](#)
  - 分布 [22](#)
  - 分类因子的选项 [24](#)
  - 估计准则 [25](#)
  - 估算边际平均值 [27](#)
  - 关联函数 [22](#)
  - 广义线性混合模型 [37](#)
  - 将变量保存到活动数据集 [27](#)
  - 模型导出 [28](#)
  - 模型规范 [25](#)
  - 模型类型 [22](#)
  - 统计 [26](#)
  - 响应 [24](#)
  - 预测变量 [24](#)
- 核心岭回归
  - 保存 [105](#)
  - 绘图 [105](#)
  - 交叉验证折数 [105](#)
  - 显示 [105](#)
- 恒等关联函数
  - 在广义估计方程中 [30](#)
  - 在广义线性模型中 [22](#)
- 互补 双对数关联函数
  - 在广义估计方程中 [30](#)
  - 在广义线性模型中 [22](#)
- 绘图
  - 在“一般对数线性分析”中 [50](#)
  - 在“Logit 对数线性分析”中 [53](#)
- 混合模型

- 混合模型 (继续)
  - 广义线性混合模型 [37](#)
  - 线性 (linear) [16](#)
- 几率比
  - 在“一般对数线性分析”中 [49](#)
- 加权预测值
  - 在“GLM 重复测量”中 [12](#)
  - 在“GLM”中 [5](#)
- 伽玛分布
  - 在广义估计方程中 [30](#)
  - 在广义线性模型中 [22](#)
- 交叉制表
  - 残差 [47](#)
- 交互项
  - 在“线性混合模型”中 [18](#)
- 刻度参数
  - 在“广义估算方程”中 [33](#)
  - 在广义线性模型中 [25](#)
- 库克距离
  - 在“GLM 重复测量”中 [12](#)
  - 在“GLM”中 [5](#)
  - 在广义线性模型中 [27](#)
- 拉格朗日乘数检验
  - 在广义线性模型中 [26](#)
- 累积 Cauchit 关联函数
  - 在广义估计方程中 [30](#)
  - 在广义线性模型中 [22](#)
- 累积 logit 关联函数
  - 在广义估计方程中 [30](#)
  - 在广义线性模型中 [22](#)
- 累积 probit 关联函数
  - 在广义估计方程中 [30](#)
  - 在广义线性模型中 [22](#)
- 累积负重对数关联函数
  - 在广义估计方程中 [30](#)
  - 在广义线性模型中 [22](#)
- 累积互补对数关联函数
  - 在广义估计方程中 [30](#)
  - 在广义线性模型中 [22](#)
- 列联表
  - 在“一般对数线性分析”中 [49](#)
- 幂关联函数
  - 在广义估计方程中 [30](#)
  - 在广义线性模型中 [22](#)
- 描述统计
  - 在“广义估计方程”中 [35](#)
  - 在“线性混合模型”中 [20](#)
  - 在广义线性模型中 [26](#)
- 模型视图
  - 广义线性混合模型 [45](#)
- 模型信息
  - 在“广义估计方程”中 [35](#)
  - 在广义线性模型中 [26](#)
- 模型选择对数线性分析
  - 定义因子范围 [48](#)
  - 命令附加功能 [49](#)
  - 模型 [48](#)
  - options [49](#)
- 内核岭
  - 程度 [104](#)
  - 单个模型 [104](#)
  - 伽玛 [104](#)
  - 模型选择 [104](#)
  - alpha [104](#)
- 内核岭 (继续)
  - coef0 [104](#)
- 内核岭回归
  - 参数 [105](#)
  - 网格参数 [105](#)
- 拟合度
  - 在“广义估计方程”中 [35](#)
  - 在广义线性模型中 [26](#)
- 逆高斯分布
  - 在广义估计方程中 [30](#)
  - 在广义线性模型中 [22](#)
- 偏差残差
  - 在广义线性模型中 [27](#)
- 频率
  - 在“模型选择对数线性分析”中 [49](#)
- 平方和
  - 在“方差成分”中 [15](#)
  - 在“线性混合模型”中 [18](#)
- 评分
  - 在“线性混合模型”中 [20](#)
- 泊松分布
  - 在广义估计方程中 [30](#)
  - 在广义线性模型中 [22](#)
- 泊松回归
  - 广义线性混合模型 [37](#)
  - 在“一般对数线性分析”中 [49](#)
- 奇数幂关联函数
  - 在广义估计方程中 [30](#)
  - 在广义线性模型中 [22](#)
- 奇异性容差
  - 在“线性混合模型”中 [20](#)
- 嵌套项
  - 在“广义估算方程”中 [33](#)
  - 在“线性混合模型”中 [18](#)
  - 在广义线性模型中 [25](#)
- 全因子模型
  - 在“方差成分”中 [14](#)
  - 在“GLM 重复测量”中 [9](#)
- 删失案例数
  - 在“寿命表”中 [53](#)
  - 在“Cox 回归”中 [71](#)
  - 在“Kaplan-Meier”中 [69](#)
- 生成类
  - 在“模型选择对数线性分析”中 [48](#)
- 生存 AFT
  - 生存分析对话框-类别变量 [60](#)
- 生存分析
  - 依时 Cox 回归 [74](#)
  - 在“寿命表”中 [53](#)
  - 在“Cox 回归”中 [71](#)
  - 在“Kaplan-Meier”中 [69](#)
  - 在内核岭回归中 [104](#)
- 生存函数
  - 在“寿命表”中 [53](#)
- 似然残差
  - 在广义线性模型中 [27](#)
- 寿命表
  - 比较因子级别 [54](#)
  - 不显示表 [54](#)
  - 风险率 [53](#)
  - 绘图 [54](#)
  - 命令附加功能 [55](#)
  - 生存函数 [53](#)
  - 生存状态变量 [59](#)

- 寿命表 (继续)
    - 示例 [53](#)
    - 统计 [53](#)
    - 因子变量 [54](#)
    - Wilcoxon (Gehan) 检验 [54](#)
  - 受限最大似然估计
    - 在“方差成分”中 [15](#)
  - 随机效应
    - 在“线性混合模型”中 [19](#), [21](#)
  - 随机效应协方差矩阵
    - 在“线性混合模型”中 [20](#)
  - 随机效应优先
    - 在“方差成分”中 [15](#)
  - 剔除残差
    - 在“GLM 重复测量”中 [12](#)
    - 在“GLM”中 [5](#)
  - 未标准化残差
    - 在“GLM 重复测量”中 [12](#)
    - 在“GLM”中 [5](#)
  - 线性混合模型
    - 保存变量 [21](#)
    - 构建项 [18](#)
    - 估计准则 [20](#)
    - 估算边际平均值 [21](#)
    - 固定效应 [18](#)
    - 交互项 [18](#)
    - 命令附加功能 [22](#)
    - 模型 [20](#)
    - 随机效应 [19](#), [21](#)
    - 协方差结构 [77](#)
  - 相关性矩阵
    - 在“广义估计方程”中 [35](#)
    - 在“线性混合模型”中 [20](#)
    - 在广义线性模型中 [26](#)
  - 向后去除
    - 残差 [47](#)
  - 协变量
    - 在“Cox 回归”中 [72](#)
  - 协方差参数检验
    - 在“线性混合模型”中 [20](#)
  - 协方差分析
    - 在“GLM 多变量”中 [1](#)
  - 协方差结构
    - 在“线性混合模型”中 [77](#)
  - 协方差矩阵
    - 在“广义估计方程”中 [35](#)
    - 在“广义估算方程”中 [33](#)
    - 在“线性混合模型”中 [20](#)
    - 在“GLM”中 [5](#)
    - 在广义线性模型中 [25](#), [26](#)
  - 一般对数线性分析
    - 保存变量 [51](#)
    - 保存预测值 [51](#)
    - 标准 [50](#)
    - 残差 [51](#)
    - 单元格计数分布 [49](#)
    - 单元格结构 [49](#)
    - 单元格协变量 [49](#)
    - 对比 [49](#)
    - 绘图 [50](#)
    - 命令附加功能 [51](#)
    - 模型规范 [50](#)
    - 显示选项 [50](#)
    - 因子 [49](#)
  - 一般对数线性分析 (继续)
    - 置信度区间 [50](#)
  - 一般可估函数
    - 在广义线性模型中 [26](#)
  - 一般线性模型
    - 广义线性混合模型 [37](#)
  - 因子
    - 在“GLM 重复测量”中 [8](#)
  - 因子级别信息
    - 在“线性混合模型”中 [20](#)
  - 预测值
    - 在“线性混合模型”中 [21](#)
    - 在“一般对数线性分析”中 [51](#)
    - 在“Logit 对数线性分析”中 [53](#)
  - 正态分布
    - 在广义估计方程中 [30](#)
    - 在广义线性模型中 [22](#)
  - 正态概率图
    - 在“模型选择对数线性分析”中 [49](#)
  - 置信度区间
    - 在“一般对数线性分析”中 [50](#)
  - 置信区间
    - 在“线性混合模型”中 [20](#)
    - 在“Logit 对数线性分析”中 [53](#)
  - 重复测量变量
    - 在“线性混合模型”中 [17](#)
  - 逐步二分
    - 在“广义估算方程”中 [33](#)
    - 在“线性混合模型”中 [20](#)
    - 在广义线性模型中 [25](#)
  - 主体变量
    - 在“线性混合模型”中 [17](#)
  - 字符串协变量
    - 在“Cox 回归”中 [72](#)
  - 纵向模型
    - 广义线性混合模型 [37](#)
  - 最大似然估计
    - 在“方差成分”中 [15](#)
  - 最小显著性差异
    - 在“GLM 多变量”中 [4](#)
    - 在“GLM 重复测量”中 [11](#)
- A**
- ANOVA
    - 在“GLM 多变量”中 [1](#)
    - 在“GLM 重复测量”中 [7](#)
- B**
- Bonferroni
    - 在“GLM 多变量”中 [4](#)
    - 在“GLM 重复测量”中 [11](#)
  - Breslow 检验
    - 在“Kaplan-Meier”中 [70](#)
- C**
- Cox 回归
    - 保存新变量 [72](#)
    - 迭代 [73](#)
    - 定义事件 [73](#)
    - 对比 [72](#)

Cox 回归 (继续)  
分类协变量 [72](#)  
风险函数 [72](#)  
绘图 [72](#)  
基线函数 [73](#)  
命令附加功能 [73](#)  
偏残差 [72](#)  
生存函数 [72](#)  
生存状态变量 [73](#)  
示例 [71](#)  
统计量 [71, 73](#)  
协变量 [71](#)  
依时协变量 [74](#)  
逐步输入和剔除 [73](#)  
字符串协变量 [72](#)  
DfBeta [72](#)

## D

disaster recovery [81](#)  
Duncan 的多范围检验  
在“GLM 重复测量”中 [11](#)  
Duncan 多重范围检验  
在“GLM 多变量”中 [4](#)  
Dunnnett 的 C  
在“GLM 重复测量”中 [11](#)  
Dunnnett 的 T3  
在“GLM 重复测量”中 [11](#)  
Dunnnett C  
在“GLM 多变量”中 [4](#)  
Dunnnett t 检验  
在“GLM 多变量”中 [4](#)  
在“GLM 重复测量”中 [11](#)  
Dunnnett T3  
在“GLM 多变量”中 [4](#)

## F

Fisher 的 LSD  
在“GLM 多变量”中 [4](#)  
在“GLM 重复测量”中 [11](#)  
Fisher 评分方法  
在“线性混合模型”中 [20](#)

## G

Gabriel 成对比较检验  
在“GLM 多变量”中 [4](#)  
Gabriel 的成对比较检验  
在“GLM 重复测量”中 [11](#)  
Games 和 Howell 成对比较检验  
在“GLM 多变量”中 [4](#)  
在“GLM 重复测量”中 [11](#)  
Gehan 检验  
在“寿命表”中 [54](#)  
GLM  
保存变量 [5](#)  
保存矩阵 [5](#)  
GLM 多变量  
概要图 [3](#)  
事后检验 [4](#)  
协变量 [1](#)  
因变量 [1](#)

GLM 多变量 (继续)  
因子 [1](#)  
GLM 重复测量  
保存变量 [12](#)  
定义因子 [8](#)  
概要图 [10](#)  
命令附加功能 [13](#)  
模型 [9](#)  
事后检验 [11](#)  
GLOR  
在“一般对数线性分析”中 [49](#)

## H

Hessian 收敛  
在“广义估算方程”中 [33](#)  
在广义线性模型中 [25](#)  
high availability [81](#)  
Hochberg GT2  
在“GLM 多变量”中 [4](#)  
在“GLM 重复测量”中 [11](#)

## K

Kaplan-Meier  
保存新变量 [70](#)  
比较因子级别 [70](#)  
定义事件 [70](#)  
绘图 [71](#)  
命令附加功能 [71](#)  
平均值和中位数生存时间 [71](#)  
生存分析表 [71](#)  
生存状态变量 [70](#)  
示例 [69](#)  
四分位数 [71](#)  
统计信息 [69, 71](#)  
因子级别的线性趋势 [70](#)

## L

L 矩阵  
在“广义估计方程”中 [35](#)  
在广义线性模型中 [26](#)  
Logistic 回归  
广义线性混合模型 [37](#)  
Logit 对数线性分析  
模型规范 [52](#)  
logit 关联函数  
在广义估计方程中 [30](#)  
在广义线性模型中 [22](#)

## M

MINQUE  
在“方差成分”中 [15](#)

## N

Newman-Keuls  
在“GLM 多变量”中 [4](#)  
在“GLM 重复测量”中 [11](#)  
Newton-Raphson 方法  
在“一般对数线性分析”中 [49](#)

Newton-Raphson 方法 (继续)  
在“Logit 对数线性分析”中 [51](#)

## P

Pearson 残差  
在“广义估计方程”中 [36](#)  
在广义线性模型中 [27](#)  
Probit 分析  
广义线性混合模型 [37](#)  
probit 关联函数  
在广义估计方程中 [30](#)  
在广义线性模型中 [22](#)

## R

R-E-G-W F  
在“GLM 多变量”中 [4](#)  
R-E-G-W F(R)  
在“GLM 重复测量”中 [11](#)  
R-E-G-W Q  
在“GLM 多变量”中 [4](#)  
R-E-G-W Q(Q)  
在“GLM 重复测量”中 [11](#)  
Ryan-Einot-Gabriel-Welsch 多范围  
在“GLM 重复测量”中 [11](#)  
Ryan-Einot-Gabriel-Welsch 多重 F  
在“GLM 多变量”中 [4](#)  
在“GLM 重复测量”中 [11](#)  
Ryan-Einot-Gabriel-Welsch 多重范围  
在“GLM 多变量”中 [4](#)

## S

S-N-K  
在“GLM 重复测量”中 [11](#)  
Scheffé 检验  
在“GLM 多变量”中 [4](#)  
在“GLM 重复测量”中 [11](#)  
Sidak t 检验  
在“GLM 多变量”中 [4](#)  
Sidak's t 检验  
在“GLM 重复测量”中 [11](#)  
Student-Newman-Keuls  
在“GLM 多变量”中 [4](#)

## T

Tamhane 的 T2  
在“GLM 重复测量”中 [11](#)  
Tamhane T2  
在“GLM 多变量”中 [4](#)  
Tarone-Ware 检验  
在“Kaplan-Meier”中 [70](#)  
Tukey 的 b 检验  
在“GLM 重复测量”中 [11](#)  
Tukey 的诚实显著性差异  
在“GLM 重复测量”中 [11](#)  
Tukey b 检验  
在“GLM 多变量”中 [4](#)  
Tukey's 真实显著性差异  
在“GLM 多变量”中 [4](#)  
Tweedie 分布

Tweedie 分布 (继续)  
在广义估计方程中 [30](#)  
在广义线性模型中 [22](#)

## W

Wald 统计  
在“一般对数线性分析”中 [49](#)  
在“Logit 对数线性分析”中 [51](#)  
Waller-Duncan t 检验  
在“GLM 多变量”中 [4](#)  
在“GLM 重复测量”中 [11](#)  
Wilcoxon 检验  
在“寿命表”中 [54](#)





