

# **IBM SPSS Regression 22**



**注释**

使用本信息及其支持的产品前，请先阅读 第 27 页的『声明』中的信息。

**产品信息**

本版本使用于 IBM® SPSS® Statistics V22.0.0 及所有后续发行版和修订版，直到在新版本中另有声明为止。

---

# 目录

## 第 1 章 为“二元 Logistic 回归”选择过程 1

## 第 2 章 Logistic 回归 . . . . . 3

Logistic 回归: 设置规则 . . . . .	4
Logistic 回归: 变量选择方法 . . . . .	4
Logistic 回归: 定义分类变量 . . . . .	4
Logistic 回归: 保存新变量 . . . . .	5
Logistic 回归: 选项 . . . . .	5
LOGISTIC REGRESSION 命令的附加功能 . . . . .	6

## 第 3 章 多项 Logistic 回归 . . . . . 7

多项 Logistic 回归 . . . . .	7
建立项 . . . . .	8
多项 Logistic 回归: 参考类别 . . . . .	8
多项 Logistic 回归: 统计 . . . . .	8
多项 Logistic 回归: 标准 . . . . .	9
多项 Logistic 回归: 选项 . . . . .	9
多项 Logistic 回归: 保存 . . . . .	10
NOMREG 命令的附加功能 . . . . .	10

## 第 4 章 Probit 分析 . . . . . 11

Probit 分析: 定义范围 . . . . .	12
Probit 分析: 选项 . . . . .	12
PROBIT 命令的附加功能 . . . . .	12

## 第 5 章 非线性回归 . . . . . 13

条件逻辑 (非线性回归) . . . . .	13
非线性回归: 参数 . . . . .	14
非线性回归常用模型 . . . . .	14

非线性回归: 损失函数 . . . . .	15
非线性回归: 参数约束 . . . . .	15
非线性回归: 保存新变量 . . . . .	16
非线性回归: 选项 . . . . .	16
解释非线性回归结果 . . . . .	16
NLR 命令的附加功能 . . . . .	16

## 第 6 章 权重估计 . . . . . 19

权重估计: 选项 . . . . .	19
WLS 命令的附加功能 . . . . .	20

## 第 7 章 二阶最小平方回归 . . . . . 21

二阶最小平方回归: 选项 . . . . .	22
2SLS 命令的附加功能 . . . . .	22

## 第 8 章 分类变量编码方案 . . . . . 23

偏差 . . . . .	23
简单 . . . . .	23
Helmert . . . . .	24
差值 . . . . .	24
多项式 . . . . .	24
重复 . . . . .	25
特殊 . . . . .	25
指示灯 . . . . .	26

## 声明 . . . . . 27

商标 . . . . .	28
--------------	----

## 索引 . . . . . 31



---

## 第 1 章 为“二元 Logistic 回归”选择过程

可使用“Logistic 回归”过程或“多项 Logistic 回归”过程拟合二元 Logistic 回归模型。每个过程中都有另一个过程所没有的选项。两者之间一个重要的理论差别是，“Logistic 回归”过程使用单个个案级别的数据生成所有预测值、残差、影响统计以及拟合优度检验，与数据的输入方式以及协变量模式数是否小于个案总数无关；而“多项 Logistic 回归”过程则通过在内部汇总个案来形成对预测变量具有相同协变量模式的子群体，而后基于这些子群体生成预测值、残差和拟合优度检验。如果所有预测变量都是分类变量或者任何连续预测变量都只取有限数量的值（每个不同协变量模式因而会有多个个案），那么子群体方法可生成有效的拟合优度检验和有意义的残差，而单个案级方法则不能。

**Logistic 回归**提供以下独特的功能：

- 模型的 Hosmer-Lemeshow 拟合优度检验
- 逐步分析
- 用于定义模型参数化的对比
- 用于分类的替代分割点
- 分类图(C)
- 将已在一组个案上拟合的模型应用到提供的一组个案
- 保存预测值、残差及影响统计

**多项 Logistic 回归**提供以下独特的功能：

- 模型拟合度优度 Pearson 和偏差卡方检验
- 用于对进行拟合优度检验的数据进行分组的子群体指定
- 按子群体列出的计数、预测计数和残差
- 对过度离散的方差估计值的修正
- 参数估计的协方差矩阵
- 参数线性组合的检验
- 嵌套模型的显式指定
- 使用差分变量拟合 1-1 匹配的条件 Logistic 回归模型



---

## 第 2 章 Logistic 回归

如果希望根据一组预测变量的值预测某种特征或结果是否存在，Logistic 回归可以提供帮助。Logistic 回归类似于线性回归模型，但更适用于因变量为二分变量的模型。Logistic 回归系数可用于估计模型中每个自变量的几率比。与判别分析相比，Logistic 回归适用于更为广泛的研究场合。

**示例。** 什么样的生活方式特征是引发冠心病 (CHD) 的危险因素？给定具有吸烟状态、饮食、锻炼、饮酒和 CHD 状态度量的患者样本，可以使用这四个生活方式变量建立一个模型，用于预测患者样本中 CHD 的存在性。然后可使用此模型为每个因子推导几率比估计值，从而获知某些信息，例如吸烟者比非吸烟者在何种程度上更易患 CHD。

**统计。** 对于每项分析，统计个案总数、选定个案数和有效个案数。对于每个分类变量，统计参数编码。对于每个步骤，统计输入或移除的变量、迭代历史记录、-2 对数似然、拟合优度、Hosmer-Lemeshow 拟合优度统计、模型卡方、改进卡方、分类表、变量间的相关性、观察组和预测概率图表以及残差卡方。对于公式中的每个变量，统计系数 ( $B$ )、 $B$  的标准误差、Wald 统计、估计的几率比 ( $\exp(B)$ )、 $\exp(B)$  的置信区间和对数似然（如果从模型中移除了项）。对于公式以外的每个变量，统计得分统计。对于每个个案，统计观察组、预测概率、预测组、残差和标准化残差。

**方法。** 可以使用变量块输入或下列任一步进法来估算模型：向前条件、向前 LR、向前 Wald、向后条件、向后 LR 或向后 Wald。

Logistic 回归数据注意事项

**数据。** 因变量应为二分变量。自变量可以是区间级别变量或分类变量；如果是分类变量，则它们应经过哑元变量或指示符编码（过程中有一个选项可自动对分类变量进行重新编码）。

**假设。** 与判别分析一样，Logistic 回归不依赖于分布假设。但是，如果预测变量呈多变量正态分布，则解会更稳定。此外，与其他形式的回归相同，预测变量中的多重共线性会导致有偏差的估计和夸大的标准误差。当组成员为真正的分类变量时，该过程最为有效；如果组成员基于连续变量的值（例如，“高智商”与“低智商”），则应考虑使用线性回归以利用由连续变量本身提供的更为丰富的信息。

**相关过程。** 使用“散点图”过程过滤数据以获得多重共线性。如果符合多变量正态性和相等方差-协方差矩阵的假设，则使用“判别分析”过程可以更快获得解。如果所有预测变量都是分类变量，则还可以使用“对数线性”过程。如果因变量是连续的，请使用“线性回归”过程。可以使用“ROC 曲线”过程绘制通过“Logistic 分析”过程保存的概率。

获取 Logistic 回归分析

1. 从菜单中选择：

分析 > 回归 > 二项 **Logistic...**

2. 选择一个二分因变量。该变量可以是数值或字符串。

3. 选择一个或多个协变量。要包含交互项，请选择交互中所涉及的所有变量，然后选择 **>a\*b>**。

要在组（块）中输入变量，请为块选择协变量，然后单击下一步以指定新块。重复此步骤，直到指定了所有的块。

或者，可以选择分析的个案。选择一个选择变量，然后单击**规则**。

---

## Logistic 回归: 设置规则

模型估计中包含由选择规则定义的个案。例如, 如果选择了一个变量和等于并指定值 5, 则估计此模型时只包含选定变量具有等于 5 的值的个案。

同时为选定的和未选定的个案生成统计和分类结果。这将提供一种机制, 以基于以前存在的数据对新个案进行分类, 或是将数据划分成训练子集和检验子集, 从而对生成的模型执行验证。

---

## Logistic 回归: 变量选择方法

方法选择允许您指定自变量将如何进入到分析中。通过使用不同的方法, 您可以根据相同的变量组构造多个回归模型。

- **输入。** 一种变量选择过程, 其中一个块中的所有变量在一个步骤中输入。
- **向前选择 (条件) (Forward Selection (Conditional)).** 逐步选择方法, 其中进入检验是基于得分统计的显著性, 移去检验是基于在条件参数估计基础上的似然比统计的概率。
- **向前选择 (似然比) (Forward Selection (Likelihood Ratio)).** 逐步选择方法, 其中进入检验是基于得分统计的显著性, 移去检验是基于在最大局部似然估计的似然比统计的概率。
- **向前选择 (Wald) (Forward Selection (Wald)).** 逐步选择方法, 其中进入检验是基于得分统计的显著性, 移去检验是基于 Wald 统计的概率。
- **向后去除 (条件) (Backward Elimination (Conditional)).** 逐步向后选择。移去检验基于在条件参数估计的似然比统计的概率。
- **向后去除 (似然比) (Backward Elimination (Likelihood Ratio)).** 逐步向后选择。移去检验基于在最大偏似然估计基础上的似然比统计的概率。
- **向后去除 (Wald) (Backward Elimination (Wald)).** 逐步向后选择。移去检验基于 Wald 统计的概率。

输出中的显著性值基于与单个模型的拟合。因此, 当使用步进法时, 显著性值通常无效。

所有被选自变量将被添加到单个回归模型中。不过, 您可以为不同的变量子集指定不同的进入方法。例如, 您可以使用逐步式选择将一个变量块输入到回归模型中, 而使用向前选择输入第二个变量块。要将第二个变量块添加到回归模型, 请单击下一个。

---

## Logistic 回归: 定义分类变量

您可以详细指定“Logistic 回归”过程处理分类变量的方式:

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

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

**更改对比。** 可用于更改对比方法。可用的对比方法有:

- **指示符。** 这些对比指示类别成员资格是否存在。参考类别在对比矩阵中表示为一排“0”。
- **简单。** 除参考类别外, 预测变量的每个类别都与参考类别相比较。
- **差分。** 除第一个类别外, 预测变量的每个类别都与前面的类别的平均效应相比较。也称为逆 Helmert 对比。
- **Helmert。** 除最后一个类别外, 预测变量的每个类别都与后面的类别的平均效应相比较。
- **重复。** 除第一个类别外, 预测变量的每个类别都与它前面的那个类别进行比较。



- **多项式。** 正交多项式对比。假设类别均匀分布。多项式对比仅适用于数值变量。
- **偏差。** 除参考类别外，预测变量的每个类别都与总体效应相比较。

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

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

---

## Logistic 回归: 保存新变量

可以将 Logistic 回归的结果保存为活动数据集中的新变量:

**预测值。** 保存由模型预测的值。可用选项有“概率”和“组成员资格”。

- **概率 (Probabilities).** 对于每个个案，保存事件发生的预测概率。输出中的表显示了任何新变量的名称和内容。“事件”是值较大的从属变量类别；例如，如果从属变量的值取 0 和 1，那么将保存类别 1 的预测概率。
- **预测组成员资格 (Predicted Group Membership).** 基于判别分数，具有最大后验概率的组。模型预测该个案所属的组。

**影响。** 保存来自测量个案对预测值影响的统计的值。可用选项有“Cook 距离”、“杠杆值”和“DfBeta”。

- **Cook 距离 (Cook's).** Cook 距离影响统计的 Logistic 回归模拟。在特定个案从回归系数的计算中排除的情况下，所有个案的残差变化幅度的测量。
- **杠杆值 (Leverage Value).** 每个观察值对模型拟合度的相对影响。
- **DfBeta(s) . beta** 值的差分是由于排除了某个特定个案而导致的回归系数的改变。为模型中的每一项（包括常数项）均计算一个值。

**残差。** 保存残差。可用的选项有“未标准化”、“Logit”、“学生化”、“标准化”和“偏差”。

- **未标准化残差 (Unstandardized Residuals).** 观察值与模型预测值之间的差。
- **Logit 残差 (Logit Residual).** 使用 logit 刻度对个案进行预测时个案的残差。Logit 是一个商，分子是残差，分母是预测概率乘以 1 与预测概率的差。
- **Student 化的残差 (Studentized Residual).** 排除了某个个案的情况下，模型离差的改变。
- **标准化残差 (Standardized Residuals).** 残差除以其标准差的估计。标准化残差也称为 Pearson 残差，它的平均值为 0，标准差为 1。
- **偏差 (Deviance).** 基于模型偏差的残差。

**将模型信息输出到 XML 文件。** 将参数估计值及其（可选）协方差导出到指定的 XML (PMML) 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。

---

## Logistic 回归: 选项

可以为 Logistic 回归分析指定选项:

**统计和图。** 允许您请求统计和图。可用选项有“分类图”、“Hosmer-Lemeshow 拟合度”、“个案的残差列表”、“估计的相关性”、“迭代历史记录”以及“exp(B) 的置信限制”。在“显示”组中选择一个选项，用于显示统计和图，可在每步中显示，或只在最后一步为最终模型显示。

- **Hosmer-Lemeshow 拟合优度统计 (Hosmer-Lemeshow goodness-of-fit statistic).** 此拟合度统计比用于 logistic 回归中所用的传统拟合度统计更稳健，特别是对于具有连续协变量的模型和使用小样本的研究。统计基于将个案分组为不同的风险度十分位数并比较每个十分位数中的已观察到的概率与期望概率。

**步进概率。** 允许您控制变量输入方程和从方程中移去的条件。可指定“输入”或“剔除”变量的条件。

- **步进概率 (Probability for Stepwise).** 如果某个变量的得分统计概率小于“输入”值，那么向模型中输入该变量；如果该概率大于“剔除”值，那么移去该变量。要覆盖缺省设置，请为“输入”和“剔除”输入正值。“输入”必须小于“剔除”。

**分类分界值。** 允许您确定分类个案的分割点。具有大于分类分界值的预测值的个案被分类为正，具有小于分类分界值的预测值的个案分类为负。要更改缺省值，请输入一个 0.01 到 0.99 之间的值。

**最大迭代次数。** 允许您更改模型在终止前的最大迭代次数。

**在模型中包含常数。** 允许您指定模型是否应包含常数项。如果禁用此选项，则常数项将等于 0。

---

## LOGISTIC REGRESSION 命令的附加功能

使用命令语法语言还可以：

- 按变量的值或变量标签逐个个案标识输出。
- 控制迭代报表的间距。可以在每  $n$  次迭代后请求参数估计值，而不是在每次迭代后都打印参数估计值。
- 更改用于终止迭代和检查冗余的条件。
- 为个案情况列表指定变量列表。
- 在处理期间将每个拆分文件组的数据保存到一个外部临时文件中，从而节省内存。

请参阅命令语法参考以获取完整的语法信息。

---

## 第 3 章 多项 Logistic 回归

如果希望基于一组预测变量的值对主体进行分类，多项 Logistic 回归非常有用。此类型的回归与 Logistic 回归类似，但更具一般性，因为因变量不限定为两个分类。

**示例。** 为更有效地宣传影片，电影厂希望预测影迷们喜欢看何种类型的电影。通过执行“多项 Logistic 回归”，电影厂可确定个人的年龄、性别以及恋爱状况对影片类型偏好的影响程度。然后，电影厂可以面向可能观看影片的人群有侧重点地开展特定影片的宣传活动的。

**统计。** 迭代历史记录、参数系数、渐近协方差和相关性矩阵、模型和偏效应的似然比检验，- 2 对数似然统计。Pearson 和偏差卡方拟合优度。Cox 和 Snell、Nagelkerke 以及 McFadden  $R^2$ 。分类：按响应类别区分的观察频率和预测频率。交叉表格：按协变量模式和响应类别区分的观察频率和预测频率（带残差）以及比例。

**方法。** 针对全因子模型或用户指定的模型拟合度多项 Logit 模型。通过迭代最大似然算法执行参数估计。

多项 Logistic 回归数据注意事项

**数据。** 因变量应为分类变量。因变量可以是因子或协变量。总的来说，因子应为分类变量，协变量应为连续变量。

**假设。** 假设任意两个类别的几率比独立于所有其他响应类别。例如，如果将新产品推向市场，则此假设认为所有其他产品的市场份额都均等地受到影响。同时，在给定的协变量模式下，假设响应独立于多项变量。

获取多项 Logistic 回归

1. 从菜单中选择：

    分析 > 回归 > 多项 Logistic...

2. 选择一个因变量。

3. 因子是可选的，可以是数值或类别。

4. 协变量是可选的，但必须是数值（如果指定）。

---

### 多项 Logistic 回归

“多项 Logistic 回归”过程缺省生成带有因子和协变量主效应的模型，但使用此对话框可以指定自定义模型，或请求逐步方式的模型选择。

**指定模型。** 主效应模型包含协变量和因子的主效应，但不包含交互效应。全因子模型包含所有主效应和所有因子与因子的交互效应。它不包含协变量交互。可以创建自定义模型以指定因子交互或协变量交互的子集，或请求逐步选择模型项。

**因子与协变量。** 列出因子与协变量。

**强制输入项。** 模型中总是包含添加到强制输入列表中的项。

**步进项。** 逐步列表中添加的项根据下列用户选择的步进法之一包含在模型中：

- **向前进入。** 此方法开始时模型中没有步进项。在每一步中将最显著的项添加到模型中，直到留在模型之外的任何步进项在添加到模型中之后都不会具有显著的统计作用。

- **向后去除。**此方法在开始时将逐步列表中指定的所有项输入到模型中。每一步从模型中移去最不显著的步进项，直到剩余所有步进项都对模型具有显著的统计作用。
- **前向逐步。**此方法从将由向前进入方法选定的模型开始。在此模型的基础上，算法交替执行模型中步进项的向后去除和模型外剩余项的向前进入。此操作持续执行，直到不再有项满足输入或移去标准。
- **向后步进。**此方法从将由向后去除方法选定的模型开始。在此模型的基础上，算法交替执行模型外剩余项的向前进入和模型中步进项的向后去除。此操作持续执行，直到不再有项满足输入或移去标准。

在模型中包含截距。允许您在模型中包含或排除截距项。

## 建立项

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。

**主效应。**为每个选定的变量创建主效应项。

**所有二阶。**创建选定变量的所有可能的双向交互。

**所有三阶。**创建选定变量的所有可能的三阶交互。

**所有四阶。**创建选定变量的所有可能的四阶交互。

**所有五阶。**创建选定变量的所有可能的五阶交互。

---

## 多项 Logistic 回归: 参考类别

缺省情况下，“多项 Logistic 回归”过程将最后一个类别作为参考类别。此对话框为您提供对参考类别的控制，以及对类别排序方式的控制。

**参考类别。**指定第一个、最后一个或自定义类别。

**类别顺序。**在升序方式中，最低值定义第一个类别，最高值定义最后一个类别。在降序方式中，最高值定义第一个类别，最低值定义最后一个类别。

---

## 多项 Logistic 回归: 统计

可为“多项 Logistic 回归”指定下列统计:

**个案处理摘要。**此表包含有关指定的分类变量的信息。

**模型。**总体模型的统计。

- **伪 R 方。**打印 Cox 和 Snell、Nagelkerke 以及 McFadden  $R^2$  统计。
- **步骤摘要。**此表汇总了步进法中每一步中进入或移去的效应。除非模型对话框中指定了步进模型，否则将不会生成此内容。
- **模型拟合度信息。**此表将拟合模型与仅截距或空模型进行比较。
- **信息准则。**此表打印 Akaike 的信息标准 (AIC) 和 Schwarz 的 Bayesian 信息标准 (BIC)。
- **单元格概率。**按协变量模式和响应类别打印观察的和期望的频率（带残差）和比例的表。
- **分类表。**打印观察响应和预测响应的表。

- **拟合度卡方统计。** 打印 Pearson 和似然比卡方统计。将为由所有因子和协变量，或由用户定义的因子和协变量子集确定的协变量模式计算统计。
- **单调性测量。** 显示包含有关协调对、非协调对和相等对的信息的表。此表中还显示 Somers 的 D、Goodman 和 Kruskal 的伽玛、Kendall 的 tau-a 以及协调索引 C。

**参数。** 与模型参数相关的统计。

- **估计。** 使用用户指定的置信度打印模型参数的估计值。
- **似然比检验。** 打印模型偏效应的似然比检验。将自动打印总体模型的检验。
- **渐近相关。** 打印参数估计相关性的矩阵。
- **渐近协方差。** 打印参数估计协方差的矩阵。

**定义子群体。** 允许您选择因子和协变量的子集，以便定义单元格概率和拟合度检验所用的协变量模式。

---

## 多项 Logistic 回归: 标准

可为“多项 Logistic 回归”指定下列标准:

**迭代。** 允许您指定希望算法循环的最大次数、步骤对分中的最大步骤数、对数似然估计和参数中的更改的收敛性容差、打印迭代算法进度的频率以及过程应在哪次迭代中开始检查完成的或准完成的数据分离。

- **对数似然估计收敛性。** 如果对数似然函数中的绝对变更小于指定值，则假定收敛性。如果值为 0，则不使用此准则。请指定一个非负值。
- **参数收敛。** 如果参数估计值中的绝对变更小于此值，则假定收敛性。如果值为 0，则不使用该标准。

**Delta。** 允许您指定一个小于 1 的非负值。此值将按协变量模式添加到响应类别交叉制表的每个空单元格中。这有助于稳定算法和避免估计值偏差。

**奇异性容差。** 允许您指定检查奇异性时所用的容差。

---

## 多项 Logistic 回归: 选项

可为“多项 Logistic 回归”指定下列选项:

**离散度量。** 允许您指定将用于修正参数协方差矩阵的估计的离散度量值。**偏差**使用偏差函数（似然比卡方）统计估计刻度值。**Pearson** 使用 Pearson 卡方统计估计刻度值。也可以指定自己的刻度值。必须是正数值。

**步进选项。** 在使用步进法建立模型时，此选项可为您提供对统计标准的控制。除非模型对话框中指定了步进模型，否则将忽略这些选项。

- **输入概率。** 这是用于变量输入的似然比统计的概率。指定的概率越大，变量就越容易进入模型。除非选择向前进入、向前步进或向后步进法，否则此准则将被忽略。
- **输入检验。** 这是用于在步进法中输入项的方法。请在似然比检验和得分检验间选择。除非选择向前进入、向前步进或向后步进法，否则此准则将被忽略。
- **删除概率。** 这是用于变量剔除的似然比统计的概率。指定的概率越大，变量就越容易保留在模型中。除非选择向后去除、向前步进或向后步进法，否则此准则将被忽略。
- **剔除检验。** 这是用于在步进法中移去项的方法。请在似然比检验和 Wald 检验间选择。除非选择向后去除、向前步进或向后步进法，否则此准则将被忽略。
- **模型中的最小步进效应。** 使用向后去除或向后步进法时，此选项指定将包含在模型中的最小项数。截距不算作模型项。

- **模型中的最大步进效应。** 使用向前进入或向前步进法时，此选项指定将包含在模型中的最大项数。截距不算作模型项。
- **分级约束项的输入和删除。** 此选项允许您选择是否对模型项的包含方式进行限制。对于要包含的任何项，层次要求属于该项的所有较低次序项必须首先位于模型中。例如，如果层次要求有效，则因子婚姻状况和性别二者必须都位于模型中，然后才能添加婚姻状况\*性别交互。三个单选按钮选项决定了协变量在确定层次过程中的角色。

---

## 多项 Logistic 回归: 保存

“保存”对话框允许您将变量保存到工作文件，并将模型信息导出到一个外部文件中。

保存的变量。 可保存以下变量:

- **估计响应概率。** 这些概率是将因子/协变量模式分类为响应类别的估计概率。估计概率的数目与响应变量的类别数目相当; 最多保存 25 个概率。
- **预测类别。** 该类别是因子/协变量模式的具有最大期望概率的响应类别。
- **预测类别概率。** 该概率是估计响应概率中的最大概率。
- **实际类别概率。** 该概率是将因子/协变量模式分类为观察类别的估计概率。

将模型信息输出到 **XML 文件**。 将参数估计值及其 (可选) 协方差导出到指定的 XML (PMML) 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。

---

## NOMREG 命令的附加功能

使用命令语法语言还可以:

- 指定因变量的参考类别。
- 包含具有用户缺失值的个案。
- 通过将原假设指定为参数的线性组合来自定义假设检验。

请参阅 *命令语法参考* 以获取完整的语法信息。



---

## 第 4 章 Probit 分析

此过程测量刺激的强度与对刺激显示出特定响应的个案比例之间的关系。如果您具有二分输出，并认为该输出受某些自变量级别的影响或是由其导致的，并且尤其适合实验数据，则此分析非常有用。使用此过程可以估计引致特定比例的响应所需的刺激强度，例如中位效应剂量。

**示例。** 新型杀虫剂对于杀灭蚂蚁的有效性如何，适用浓度多大？可以执行一项实验，对蚂蚁样本施用不同浓度的杀虫剂，然后记录杀灭的蚂蚁数量以及被施用杀虫剂的蚂蚁数量。通过对这些数据应用 Probit 分析，可以确定浓度和杀灭效力之间的关系紧密度，并且可以确定在希望确保杀灭一定比例（例如 95%）的蚂蚁时杀虫剂的适当浓度。

**统计。** 回归系数和标准误差、截距和标准误差、Pearson 拟合优度卡方、观察的和期望的频率以及自变量有效级别的置信区间。绘图：已转换响应图。

此过程使用 Gill、Murray、Saunders 和 Wright 提出并由其在 NPSOL<sup>®</sup> 中实现的算法来估计模型参数。

### Probit 分析数据注意事项

**数据。** 对于自变量的每个值（或多个自变量的每个值组合），响应变量应为具有显示相应响应的值的个案数。观察变量总数应为自变量具有这些值的个案的总数。因子变量应是以整数编码的分类变量。

**假设。** 观察值应是独立的。如果自变量值的数量与观察值的数量相比过多（在某项观察研究中可能遇到这样的情况），则卡方统计和拟合优度统计可能无效。

**相关过程。** Probit 分析与 Logistic 回归紧密相关；实际上，如果选择 Logit 转换，则此过程最终计算的是 Logistic 回归。总的来说，Probit 分析适用于设计的实验，而 Logistic 回归更适用于观察研究。输出中的差异反映了这些不同的侧重方面。Probit 分析过程报告不同响应频率下有效值的估计值（包括中位效应剂量），而 Logistic 回归过程报告自变量几率比的估计值。

### 获取 Probit 分析

1. 从菜单中选择：

分析 > 回归 > **Probit...**

2. 选择一个响应频率变量。此变量表示对检验刺激表现出响应的个案数。此变量的值不能为负。
3. 选择观察变量总数。此变量表示应用刺激的个案数。此变量的值不能为负，并且不能少于每个个案的响应频率变量的值。

根据需要，可以选择“因子”变量。如果选择此变量，请单击**定义范围**来定义组。

4. 选择一个或多个协变量。此变量包含应用到每个观察的刺激级别。如果要转换协变量，请从“转换”下拉列表表中选择一个转换。如果不应用任何转换，并且有一个控制组，则分析中将包含该控制组。
5. 选择 **Probit** 或 **Logit** 模型。
  - *Probit 模型 (Probit Model)*. 对响应比例应用 probit 转换（累积标准正态分布函数的逆函数）。
  - *Logit 模型 (Logit Model)*. 对响应比例应用 logit（对数几率）转换。

---

## Probit 分析: 定义范围

在此对话框中可以指定将分析的因子变量的水平。因子级别必须编码为连续整数，过程将对指定范围中的所有水平进行分析。

---

## Probit 分析: 选项

您可以为 Probit 分析指定选项:

**统计。** 使您能够请求获取以下可选统计信息: 频率、相对中位数、平行检验和基准置信区间。

- **相对中位数 (Relative Median Potency).** 显示每对因子级别的中位数强度比。还显示每个相对中位数的 95% 置信限制。如果您没有因子变量或具有多个协变量，那么相关中位数力不可用。
- **平行检验 (Parallelism Test).** 对所有因子级别具有共同的斜率这一假设的检验。
- **信仰置信区间 (Fiducial Confidence Intervals).** 生成确定的响应概率所必需的代理用量的置信区间。

如果选择了多个协变量，则信仰置信区间和相对中位数不可用。只有在选择了因子变量的情况下，相对中位数和平行检验才可用。

**自然响应频率。** 允许您指定自然响应频率，即使在没有刺激的情况下也可以。可用选项有“无”、“从数据中计算”和“值”。

- **从数据中计算 (Calculate from Data).** 根据样本数据估计自然响应频率。数据应包含代表控制级别的个案，而该级别的协变量值为 0。Probit 使用该控制级别的响应比例来估计自然响应率以作为初始值。
- **值 (Value).** 在模型中设置自然响应率（当您预先知道自然响应率时，选择此项）。输入自然响应比例（该比例必须小于 1）。例如，如果当激励为 0 时响应在 10% 的时间里发生，那么输入 0.10。

**标准。** 允许您控制迭代参数估计算法的参数。可以覆盖“最大迭代次数”、“步骤限制”和“最优性容差”的缺省值。

---

## PROBIT 命令的附加功能

使用命令语法语言还可以:

- 对 Probit 和 Logit 模型请求分析。
- 控制对缺失值的处理。
- 通过 10 和自然对数底数之外的底数转换协变量。

请参阅命令语法参考以获取完整的语法信息。



---

## 第 5 章 非线性回归

非线性回归是寻找因变量和一组自变量之间关系的非线性模型的方法。与限制为估计线性模型的传统线性回归不同，非线性回归可估计自变量和因变量之间具有任意关系的模型。这是通过使用迭代估计算法实现的。请注意，对于形式为  $Y = A + BX^{**2}$  的简单多项式模型，此过程是不必要的。通过定义  $W = X^{**2}$ ，可得到一个简单的线性模型  $Y = A + BW$ ，使用传统方法如“线性回归”过程便可以估计这一模型。

**示例。** 是否可根据时间来预测人口？散点图显示在人口和时间之间似乎有很紧密的关系，但是此关系是非线性的，因此需要使用“非线性回归”过程的特殊估计方法。通过建立适当的方程，如 Logistic 人口增长模型，我们可以获得对模型的适当估计，从而对没有实际进行测量的时间点的人口进行预测。

**统计。** 对于每次迭代，统计参数估计值和残差平方和。对于每个模型，统计回归的平方和、残差、未修正总数和修正总数、参数估计值、渐近标准误差以及参数估计值的渐近相关性矩阵。

**注：**受约束的非线性回归使用 Gill、Murray、Saunders 和 Wright 提出并由其在 NPSOL<sup>®</sup> 中实现的算法来估计模型参数。

非线性回归数据注意事项

**数据。** 因变量和自变量必须是定量的。分类变量（例如宗教、专业或居住地）需要重新编码为二分类（哑元）变量或其他类型的对比变量。

**假设。** 只有在指定的函数能够准确描述因变量和自变量的关系的情况下，结果才是有效的。此外，选择合适的起始值也非常重要。即使指定了模型的正确函数形式，如果使用不合适的起始值，模型也可能不收敛性，或者可能得到局部最优的解，而不是全局最优的解。

**相关过程。** 很多在一开始呈现为非线性的模型都可以转换为线性模型，从而使用“线性回归”过程进行分析。如果不确定什么样的模型合适，可以使用“曲线估计”过程帮助确定数据中有用的函数关系。

获取非线性回归分析

1. 从菜单中选择：

    分析 > 回归 > 非线性...

2. 从活动数据集中的变量列表选择一个数值因变量。

3. 要建立模型表达式，请在“模型”字段中输入表达式，或将成分（变量、参数和函数）粘贴到此字段中。

4. 单击**参数**标识模型中的参数。

必须在单个模型语句中使用条件逻辑指定分段模型（在域的不同部分采用不同形式的模型）。

---

### 条件逻辑（非线性回归）

可以使用条件逻辑指定分段模型。要在模型表达式或损失函数中使用条件逻辑，需要形成一系列项的总和，每个条件对应一个项。每个项由一个逻辑表达式（放在括号中）乘以一个在此逻辑表达式为真时应得到的结果表达式构成。

例如，考虑一个分段模型，它在  $X \leq 0$  时等于 0，在  $0 < X < 1$  时等于  $X$ ，在  $X \geq 1$  时等于 1。该表达式为：

$(X \leq 0) * 0 + (X > 0 \ \& \ X < 1) * X + (X \geq 1) * 1.$

括号中的逻辑表达式都计算为 1（真）或 0（假）。因此：

如果  $X \leq 0$ ，则上面的表达式简化为  $1 * 0 + 0 * X + 0 * 1 = 0$ 。

如果  $0 < X < 1$ ，则简化为  $0 * 0 + 1 * X + 0 * 1 = X$ 。

如果  $X \geq 1$ ，则简化为  $0 * 0 + 0 * X + 1 * 1 = 1$ 。

通过替换不同的逻辑表达式和结果表达式，可以很容易地建立更为复杂的示例。请记住，双向不等式（如  $0 < X < 1$ ）必须写为复合表达式（如  $X > 0 \ \& \ X < 1$ ）。

可以在逻辑表达式中使用字符串变量：

```
(city='New York')*costliv + (city='Des Moines')*0.59*costliv
```

对于 New York 的居民，将生成一个表达式（变量 *costliv* 的值），而对 Des Moines 的居民生成另一个表达式（前面的值的 59%）。字符串常数必须包含在引号或撇号中，如此处所示。

---

## 非线性回归：参数

参数是“非线性回归”过程所估计的模型的一部分。参数可以是可加常数、可乘系数、指数或函数计算中使用的值。定义的所有参数都将显示（带有其初始值）在主对话框中的“参数”列表中。

**名称。** 必须为每个参数指定名称。此名称必须是有效的变量名，且必须是在主对话框的模型表达式中使用的名称。

**初始值。** 允许您为参数指定一个初始值，与期望的最终解越接近越好。不合适的初始值可能导致收敛性失败或者导致局部（而不是全局）解的收敛性，或者在物理上是不可行的。

**使用上一分析的起始值。** 如果已经从此对话框运行过非线性回归，则可以选中此选项从上一次运行中的参数值中获取参数起始值。这允许您在算法缓慢收敛性时继续搜索。（这些起始值仍将显示在主对话框的“参数”列表中）。

注：在会话的其余部分，此项选择在对话框中持续存在。如果更改模型，请确保取消选择该选项。

---

## 非线性回归常用模型

下表提供了很多已发布的非线性回归模型的示例模型语法。随机选择的模型不太可能很好地拟合数据。参数的适当起始值是必要的，有些模型还要求使用约束才能实现收敛性。

表 1. 示例模型语法

名称	模型表达式
渐近回归	$b1 + b2 * \exp(b3 * x)$
渐近回归	$b1 - (b2 * (b3 ** x))$
密度(D)	$(b1 + b2 * x) ** (-1 / b3)$
Gauss	$b1 * (1 - b3 * \exp(-b2 * x ** 2))$
Gompertz	$b1 * \exp(-b2 * \exp(-b3 * x))$
Johnson-Schumacher	$b1 * \exp(-b2 / (x + b3))$
对数修改	$(b1 + b3 * x) ** b2$
对数 Logistic	$b1 - \ln(1 + b2 * \exp(-b3 * x))$

表 1. 示例模型语法 (续)

名称	模型表达式
Metcherlich 的收益递减规律	$b1 + b2 * \exp(-b3 * x)$
Michaelis Menten	$b1 * x / (x + b2)$
Morgan-Mercer-Florin	$(b1 * b2 + b3 * x ** b4) / (b2 + x ** b4)$
Peal-Reed	$b1 / (1+ b2 * \exp(-(b3 * x + b4 * x **2 + b5 * x ** 3)))$
三次比	$(b1 + b2 * x + b3 * x ** 2 + b4 * x ** 3) / (b5 * x ** 3)$
四次比	$(b1 + b2 * x + b3 * x ** 2) / (b4 * x ** 2)$
Richards	$b1 / ((1 + b3 * \exp(-b2 * x)) ** (1 / b4))$
Verhulst	$b1 / (1 + b3 * \exp(-b2 * x))$
Von Bertalanffy	$(b1 ** (1 - b4) - b2 * \exp(-b3 * x)) ** (1 / (1 - b4))$
韦伯	$b1 - b2 * \exp(-b3 * x ** b4)$
产量密度	$(b1 + b2 * x + b3 * x ** 2) ** (-1)$

## 非线性回归: 损失函数

非线性回归中的损失函数是由算法最小化的函数。请选择残差的平方和以最小化残差的平方和, 或选择用户定义的损失函数以最小化不同的函数。

如果选择用户定义损失函数, 则必须定义其总和 (针对所有个案) 应由所选参数值最小化的损失函数。

- 多数损失函数都涉及到特殊变量 *RESID\_*, 它代表残差。(缺省残差平方和损失函数可显式输入为 *RESID\_\*\*2*。) 如果需要在损失函数中使用预测值, 则它等于因变量减去残差。
- 可以使用条件逻辑指定条件损失函数。

您既可以在“用户定义损失函数”字段中键入一个表达式, 也可以将表达式的成分粘贴到该字段中。字符串常数必须包含在引号或撇号中, 数字常数必须以美式格式键入, 并用句点作为小数分隔符。

## 非线性回归: 参数约束

约束是在对解的迭代搜索过程中对参数所允许的值的限制。线性表达式是在步骤执行前计算的, 因此可使用线性约束来避免可能导致溢出的步骤。非线性表达式则是在步骤执行后计算的。

每个等式或不等式都要求使用下列元素:

- 至少包含模型中一个参数的表达式。键入表达式或使用小键盘, 通过小键盘可以将数字、运算符或括号粘贴到该表达式中。您既可以与表达式的其余部分一起键入所需的参数, 也可以从左侧的“参数”列表中进行粘贴。不能在约束中使用一般变量。
- 下列三个逻辑运算符中的一个,  $\leq$ 、 $=$  或  $\geq$ 。
- 数字常数, 使用逻辑运算符与表达式相比较。键入常数。数字常数必须以美式格式键入, 并使用句点作为小数分隔符。

---

## 非线性回归: 保存新变量

您可以将若干新变量保存到活动数据文件中。可用的选项有“预测值”、“残差”、“导数”和“损失函数的值”。在后续的分析中可使用这些变量, 以便检验模型的拟合度或确定有问题的个案。

- 残差 (*Residuals*). 以变量名 `resid` 保存残差。
- 预测值 (*Predicted Values*). 以变量名称 `pred_` 来保存预测值。
- 导数 (*Derivatives*). 为每个模型参数保存一个导数。通过向参数名称的前六个字符加前缀“d.”来创建导数名称。
- 损失函数值 (*Loss Function Values*). 如果您指定了自己的损失函数, 那么此选项可用。向损失函数的值分配变量名 `loss_`。

---

## 非线性回归: 选项

通过选项可以控制非线性回归分析的各个方面:

*Bootstrap 估计 (Bootstrap Estimates)*. 使用从原始数据集重复抽样来估计统计的标准误差的方法。其做法是: 抽取 (有放回取样) 样本量与原始数据集样本量相同的许多样本。为这些样本中的每一个都估计一个非线性方程。然后计算每个参数估计的标准误差, 作为自引导估计的标准差。从原始数据得到的参数估计值用作每个 Bootstrap 样本的起始值。这需要序列二次规划算法。

**估计方法。** 允许您选择一种估计方法 (如果可能)。(此对话框或其他对话框中的某些选项要求使用序列二次编程算法。) 可用的方法包括序列二次编程和 *Levenberg-Marquardt*。

- *序列二次编程 (Sequential Quadratic Programming)*. 此方法可用于约束和不约束模型。如果您指定约束模型、用户定义的损失函数或 *bootstrap*, 那么自动使用序列二次编程。您可以为“最大迭代次数”和“步长限制”输入新值, 并且可以在“最优性容差”、“函数精确度”和“无限步长”的下拉列表中更改选择。
- *Levenberg-Marquardt*. 这是非约束模型的缺省算法。如果您指定约束模型、用户定义的损失函数或自引导, 那么 *Levenberg-Marquardt* 方法不可用。您可以为“最大迭代次数”输入新值, 并且可以更改在“平方和收敛性”和“参数收敛”的下拉列表中的选择。

---

## 解释非线性回归结果

非线性回归问题经常带来计算方面的困难:

- 参数初始值的选择会影响收敛性。尽量选择合理的初始值, 并尽可能选择接近期望的最终解的初始值。
- 对于特定的问题, 一种算法的性能有时会优于另一种算法。在“选项”对话框中, 选择其他算法 (如果有)。(如果指定损失函数或特定类型的约束, 则将无法使用 *Levenberg-Marquardt* 算法。)
- 如果仅仅是因为达到迭代的最大次数而使迭代停止, 则“最终”模型可能不是一个好的解。请在“参数”对话框中选择使用先前分析中的起始值以继续迭代过程, 或者选择不同的初始值, 后一种做法更好。
- 要求对大数据值执行求幂运算或指数运算的模型可能导致溢出或下溢 (数字太大或太小, 计算机无法表示)。有时候, 通过选择适当的初始值, 或对参数施加约束, 可以避免这些问题。

---

## NLR 命令的附加功能

使用命令语法语言还可以:

- 命名一个文件, 以从其读取参数估计值的初始值。
- 指定多个模型语句和损失函数。这方便了分段模型的指定。
- 提供自己的导数, 而不是使用程序计算出的导数。
- 指定要生成的 *bootstrap* 样本的数目。

- 指定其他迭代标准，包括设置导数检查的临界值，定义残差和导数之间的相关性的收敛性准则。

通过 CNLR（受约束的非线性回归）命令的其他标准可以：

- 指定每个主迭代中允许的最大辅迭代次数。
- 设置导数检查的临界值。
- 设置步长限制。
- 指定一个崩溃容差，用于确定初始值是否位于指定边界内。

请参阅命令语法参考以获取完整的语法信息。



---

## 第 6 章 权重估计

标准线性回归模型假设方差在待研究总体中是恒定的。当方差不恒定时（例如某属性具有高属性值的个案与该属性具有低属性值的个案相比具有更大的可变性时），使用普通最小平方法 (OLS) 的线性回归不再提供最优模型估计。如果可以从另一个变量预测可变性差异，则“权重估计”过程可以使用加权最小平方 (WLS) 计算线性回归模型的系数，这样在确定回归系数时，将对更精确的观察值（即可变性小的观察值）赋予更高的权重。“权重估计”过程检验一系列权重转换，并指示给出最佳数据拟合的转换。

**示例。** 通货膨胀和失业对股票价格变化有何影响？由于高股值股票通常比低股值股票表现出更大的可变性，因此普通最小二乘法不会得出最优估计。权重估计允许您在计算线性模型时考虑股价对股价变化可变性的影响。

**统计。** 所检验的权重源变量的每个幂的对数似然估计值、复  $R$ 、 $R$  方、调整  $R$  方、WLS 模型的 ANOVA 表、未标准化和标准化参数估计值以及 WLS 模型的对数似然估计。

### 权重估计数据注意事项

**数据。** 因变量和自变量必须是定量的。分类变量（例如宗教、专业或居住地）需要重新编码为二分类（哑元）变量或其他类型的对比变量。权重变量应为定量变量，并与因变量的可变性相关。

**假设。** 对于自变量的每个值，因变量必须呈正态分布。因变量和每个自变量之间的关系应是线性的，且所有观察值应是独立的。因变量的方差对于自变量的不同级别可能不同，但是必须能够根据权重变量预测此差异。

**相关过程。** 用“探索”过程可以将数据显示到屏幕上。“探索”提供了正态性和方差同质性检验，并提供了图形化显示。如果对于自变量的不同级别，因变量都显示出具有相等的方差，则可使用“线性回归”过程。如果您的数据违反了某个假设（例如正态性假设），则尝试转换它们。如果您的数据非线性相关，且转换也没有帮助，则使用“曲线估计”过程中的备用模型。如果因变量是二分变量（例如指示特定销售是否已完成，或者某商品是否有缺陷），则请使用“Logistic 回归”过程。如果因变量是已审查变量（例如外科手术之后的存活时间），则使用 Advanced Statistics 选项中提供的“寿命表”、“Kaplan-Meier”或“Cox 回归”。如果您的数据不独立（例如，如果您在多个条件下观察同一个人），请使用 Advanced Statistics 选项中的“重复测量”过程。

### 获取权重估计分析

1. 从菜单中选择：

分析 > 回归 > 权重估计...

2. 选择一个因变量。

3. 选择一个或多个自变量。

4. 选择导致异方差性的变量作为权重变量。

- **权重变量 (Weight Variable).** 以此变量取幂后的倒数对数据进行加权。为指定范围的每个幂值计算回归方程，且回归方程指示使对数似然函数最大的幂。
- **幂范围 (Power Range).** 与权重变量结合使用以计算权重。将拟合多元回归方程，分别对应幂范围中的每个值。在“幂范围”文本框和“到”文本框中输入的值必须在 -6.5 和 7.5 之间，含 -6.5 和 7.5。幂的值范围从低值到高值，增量由指定的值确定。幂范围中值的总数不能超过 150。

---

## 权重估计：选项

您可以为权重估计分析指定选项：

将最优权重另存为新变量。将权重变量添加到活动文件。此变量称为  $WGT_n$ ，其中  $n$  是为赋予变量唯一名称而选择的数字。

显示 ANOVA 和估计。允许您控制如何在输出中显示统计。可用选项有“对于最优幂”和“为每个幂值”。

---

## WLS 命令的附加功能

使用命令语法语言还可以：

- 为幂提供单一值。
- 指定一个幂值列表，或混合值范围与幂值列表。

请参阅命令语法参考以获取完整的语法信息。



---

## 第 7 章 二阶最小平方回归

标准线性回归模型假设因变量中的误差与自变量不相关。如果不是这种情况（例如，变量间的关系是双向的），则使用普通最小平方方法 (OLS) 的线性回归不再提供最佳模型估计。二阶最小平方回归使用与误差项不相关的工具变量来计算有问题的预测变量的估计值（一阶），然后使用计算出的值来估计因变量的线性回归模型（二阶）。由于所计算的值基于与误差不相关的变量，所以二阶模型的结果是最优的。

**示例。** 对某种商品的需求是否与其价格和消费者的收入相关？此模型中的困难之处是，价格和需求互相具有倒数作用关系。即，价格可以影响需求，而需求也可以影响价格。二阶最小平方回归模型可能使用消费者的收入和延迟的价格，计算与需求中的测量误差无关的价格代理。此代理可替换原先指定的模型中的价格本身，然后对代理进行估计。

**统计。** 对于每个模型，统计标准和非标准回归系数、复  $R$ 、 $R^2$ 、调整  $R^2$ 、估计的标准误差、方差分析表、预测值和残差。此外，还有用于每个回归系数的 95% 的置信区间，以及参数估计的相关性和协方差矩阵。

二阶最小平方回归数据注意事项

**数据。** 因变量和自变量必须是定量的。分类变量（例如宗教、专业或居住地）需要重新编码为二分类（哑元）变量或其他类型的对比变量。内生解释变量应是定量变量（非分类变量）。

**假设。** 对于自变量的每个值，因变量必须呈正态分布。对于自变量的所有值，因变量分布的方差必须是恒定的。因变量和每个自变量之间的关系应为线性关系。

**相关过程。** 如果确信没有任何预测变量与因变量中的误差相关，则可使用“线性回归”过程。如果您的数据违反了假设之一（例如，正态性假设或恒定方差假设），则尝试转换数据。如果您的数据非线性相关，且转换也没有帮助，则使用“曲线估计”过程中的备用模型。如果因变量是二分变量，例如指示特定的销售是否已完成，则请使用“Logistic 回归”过程。如果您的数据不独立（例如，如果您在多个条件下观察同一个人），请使用 Advanced Models 选项中的“重复测量”过程。

获取二阶最小平方回归分析

1. 从菜单中选择:

分析 > 回归 > 二阶最小平方...

2. 选择一个因变量。

3. 选择一个或多个解释（预测）变量。

4. 选择一个或多个工具变量。

- **工具变量 (Instrumental).** 这些是在二阶最小二乘分析中的第一阶段中，用于为内生变量计算预测值的变量。相同的变量可以同时出现在“解释变量”和“工具变量”列表框中。工具变量的个数至少不能少于解释变量的个数。如果所有列出的解释变量和工具变量均相同，那么结果与“线性回归”过程的结果相同。

未指定为工具变量的解释变量被视为内生变量。通常情况下，“解释变量”列表中的所有外生变量同时也被指定为工具变量。

---

## 二阶最小平方回归: 选项

您可以为分析选择以下选项:

**保存新变量。** 允许您向活动文件中添加新变量。可用选项为“预测”和“残差”。

**显示参数估计的协方差矩阵。** 允许您打印参数估计的协方差矩阵。

---

## 2SLS 命令的附加功能

命令语法语言还允许同时估计多个方程式。请参见 [命令语法参考](#) 以获取完整的语法信息。

---

## 第 8 章 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)
```

---

## Helmert

**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)
```

---

## 差值

**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)  平均值计算的权重
(3 -1 -1 -1)  将第一项与第二到第四项进行对比
(0 2 -1 -1)  将第二项与第三和第四项进行对比
(0 0 1 -1)  将第三项与第四项进行对比

```

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 )
```

除平均值行之外的每行的和都为 0。每对非联合行的积的和也为 0：

```

第二行和第三行: (3)(0) + (-1)(2) + (-1)(-1) + (-1)(-1) = 0
第二行和第四行: (3)(0) + (-1)(0) + (-1)(1) + (-1)(-1) = 0
第三行和第四行: (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。

---

## 指示灯

**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  $k-1$ 。对于所有的  $k-1$  个变量，参考类别中的个案数的代码为 0。第  $i$  类中第  $i$  个指示符变量中的个案代码为 1，其余所有指示符变量中的个案代码均为 0。

---

## 声明

本信息是为在美国提供的产品和服务编写的。

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

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

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

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

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

本条款不适用英国或任何这样的条款与当地法律不一致的国家或地区：International Business Machines Corporation“按现状”提供本出版物，不附有任何种类的（无论是明示的还是暗含的）保证，包括但不限于暗含的有关非侵权、适销和适用于某种特定用途的保证。有些州/省不允许针对某些交易的明示或暗示免责条款，因此本声明可能不适用于您。

本信息可能含有技术误差或排版错误。此处的信息会定期进行更改；这些更改会体现在本出版物的新版本中。IBM 可以随时对本资料中描述的产品和/或程序进行改进和/或更改，而不另行通知。

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

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

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

IBM Software Group  
ATTN: Licensing

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

此类信息的提供应遵照相关条款和条件，其中包括在某些情况下支付适当费用。

本资料中描述的许可程序及其所有可用的许可资料均由 IBM 依据 IBM 客户协议、IBM 国际软件许可协议或任何同等协议中的条款提供。

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

非 IBM 产品的相关信息来自这些产品的供应商，及其发布的公告或其他公开来源。IBM 没有对这些产品进行测试，也无法确认其性能的精确性、兼容性或任何其他关于非 IBM 产品的声明。有关非 IBM 产品性能的问题应当向这些产品的供应商提出。

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

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

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：

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

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

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

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

---

## 商标

IBM、IBM 徽标和 ibm.com 是 International Business Machines Corp., 在全球许多管辖区域注册的商标或注册商标。其他产品和服务名称可能是 IBM 或其他公司的商标。最新的 IBM 商标列表可从 Web 站点 [www.ibm.com/legal/copytrade.shtml](http://www.ibm.com/legal/copytrade.shtml) 上的『版权和商标信息』部分获取。

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

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

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

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

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



Java 和所有基于 Java 的商标和徽标是 Oracle 及/或其分支机构的商标和注册商标。



# 索引

## [ B ]

步骤对分  
在“多项 Logistic 回归”中 9

## [ C ]

参考类别  
在“多项 Logistic 回归”中 8  
参数估计值  
在“多项 Logistic 回归”中 8  
参数约束  
在“非线性回归”中 15  
产量密度模型  
在“非线性回归”中 14  
常数项  
在“线性回归”中 5

## [ D ]

单元格概率表  
在“多项 Logistic 回归”中 8  
迭代  
在“多项 Logistic 回归”中 9  
在“Logistic 回归”中 5  
在“Probit 分析”中 12  
迭代历史记录  
在“多项 Logistic 回归”中 9  
对比  
在“Logistic 回归”中 4  
对数似然估计  
在权重估计中 19  
在“多项 Logistic 回归”中 8  
对数修改模型  
在“非线性回归”中 14  
多项 Logistic 回归 7, 8  
保存 10  
标准 9  
参考类别 8  
导出模型信息 10  
命令附加功能 10  
模型 7  
统计 8

## [ E ]

二阶最小平方回归 21  
保存新变量 22  
参数协方差 22  
工具变量 21

二阶最小平方回归 (续)  
命令附加功能 22  
示例 21  
统计 21  
二元 Logistic 回归 1

## [ F ]

非线性回归 13  
保存新变量 16  
参数 14  
参数约束 15  
残差 16  
常用非线性模型 14  
导数 16  
分段模型 13  
估计方法 16  
解释结果 16  
命令附加功能 16  
起始值 14  
示例 13  
损失函数 15  
条件逻辑 13  
统计 13  
序列二次规划 16  
引导估计 16  
预测值 16  
Levenberg-Marquardt 算法 16  
非线性模型  
在“非线性回归”中 14  
分类  
在“多项 Logistic 回归”中 7  
分类表  
在“多项 Logistic 回归”中 8  
分类协变量 4  
分离  
在“多项 Logistic 回归”中 9

## [ G ]

杠杆值  
在“Logistic 回归”中 5

## [ J ]

奇异性  
在“多项 Logistic 回归”中 9  
渐近回归  
在“非线性回归”中 14

截距  
包含或排除 7  
具有零观察值的单元格  
在“多项 Logistic 回归”中 9

## [ L ]

离差刻度值  
在“多项 Logistic 回归”中 9

## [ M ]

密度模型  
在“非线性回归”中 14

## [ N ]

拟合优度  
在“多项 Logistic 回归”中 8

## [ P ]

偏差函数  
用于估计离差刻度值 9  
平行检验  
在“Probit 分析”中 12

## [ Q ]

权重估计 19  
迭代历史记录 19  
对数似然估计 19  
将最佳权重另存为新变量 19  
命令附加功能 20  
示例 19  
统计 19  
显示 ANOVA 和估计 19  
全因子模型  
在“多项 Logistic 回归”中 7

## [ S ]

三次模型比  
在“非线性回归”中 14  
收敛性准则  
在“多项 Logistic 回归”中 9  
四次模型比  
在“非线性回归”中 14

似然比  
拟合优度 8  
用于估计离差刻度值 9

## [ X ]

线性回归  
二阶最小平方回归 21  
权重估计 19  
相对中位数  
在“Probit 分析”中 12  
相关性矩阵  
在“多项 Logistic 回归”中 8  
向后去除  
在“Logistic 回归”中 4  
向前选择  
在“Logistic 回归”中 4  
协变量  
在“Logistic 回归”中 4  
协方差矩阵  
在“多项 Logistic 回归”中 8  
信仰置信区间  
在“Probit 分析”中 12

## [ Y ]

约束回归  
在“非线性回归”中 15

## [ Z ]

置信区间  
在“多项 Logistic 回归”中 8  
逐步式选择  
在“多项 Logistic 回归”中 7  
在“Logistic 回归”中 4  
主效应模型  
在“多项 Logistic 回归”中 7  
自定义模型  
在“多项 Logistic 回归”中 7  
字符串协变量  
在“Logistic 回归”中 4

## C

Cook 距离  
在“Logistic 回归”中 5  
Cox 和 Snell R 方  
在“多项 Logistic 回归”中 8

## D

delta  
作为对具有零观察值的单元格的修正 9  
DfBeta  
在“Logistic 回归”中 5

## G

Gauss 模型  
在“非线性回归”中 14  
Gompertz 模型  
在“非线性回归”中 14

## H

Hosmer-Lemeshow 拟合优度统计  
在“Logistic 回归”中 5

## J

Johnson-Schumacher 模型  
在“非线性回归”中 14

## L

Logistic 回归 3  
保存新变量 5  
变量选择方法 4  
残差 5  
常数项 5  
迭代 5  
定义选择规则 4  
对比 4  
二元 1  
分类分界值 5  
分类协变量 4  
命令附加功能 6  
设置规则 4  
示例 3  
统计 3  
统计和图 5  
系数 3  
显示选项 5  
影响测量 5  
预测值 5  
逐步概率 5  
字符串协变量 4  
Hosmer-Lemeshow 拟合优度统计 5

## M

McFadden R 方  
在“多项 Logistic 回归”中 8

Metcherlich 的收益递减规律  
在“非线性回归”中 14  
Michaelis Menten 模型  
在“非线性回归”中 14  
Morgan-Mercer-Florin 模型  
在“非线性回归”中 14

## N

Nagelkerke R 方  
在“多项 Logistic 回归”中 8

## P

Peal-Reed 模型  
在“非线性回归”中 14  
Pearson 卡方  
拟合优度 8  
用于估计离差刻度值 9  
Probit 分析  
标准 12  
迭代 12  
定义范围 12  
命令附加功能 12  
平行检验 12  
示例 11  
统计 11, 12  
相对中位数 12  
信仰置信区间 12  
自然响应频率 12

## R

Richards 模型  
在“非线性回归”中 14

## V

Verhulst 模型  
在“非线性回归”中 14  
Von Bertalanffy 模型  
在“非线性回归”中 14

## W

Weibull 模型  
在“非线性回归”中 14





Printed in China