IBM SPSS Direct Marketing 20

注意：使用本信息及其支持的产品之前，请阅读注意事项第 97 页码下的一般信息。

此版本适用于 IBM® SPSS® Statistics 20 及所有后续发布和修订，除非在新版本中另有说明。

Adobe 产品屏幕截图重印已获得 Adobe Systems Incorporated 的许可。

Microsoft 产品屏幕截图重印已获得 Microsoft Corporation 的许可。

前言

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

关于 IBM Business Analytics

IBM Business Analytics 软件提供决策者赖以提高业务绩效的完整、一致和准确的信息。包括业务智能、预测分析、财务状况和战略管理以及分析应用程序在内的一整套产品组合让您即刻、清楚地了解当前绩效并依此采取行动，以及能够预测未来的成果。结合丰富的行业解决方案、被证明的实践经验和专业的服务，无论公司规模大小，都能促使其获得最高的产能、自信自觉地做出决定并得到更好的成绩。

作为产品组合的一部分，IBM SPSS Predictive Analytics 软件帮助公司预测未来实践并采取积极行动，促使其获得更好的业务成果。全世界的商业政府和学术客户依赖 IBM SPSS 技术，因其具有竞争力的优势，能够吸引、留住和发展客户，同时减少欺诈和减轻风险。通过将 IBM SPSS 软件融入日常运营中，公司成为具有预测性的企业，能够引导和自觉做出决策，以满足业务目标，实现可观的竞争优势。欲知更多信息或联系代表，请访问 http://www.ibm.com/spss。

技术支持

我们提供有“技术支持”以维护客户。客户可就 IBM Corp. 产品使用或某一受支持硬件环境的安装帮助寻求技术支持。如要联系技术支持，请访问 IBM Corp. 网站，网址为 http://www.ibm.com/support。在请求协助时，请准备好您和您组织的 ID 以及支持协议。

针对学生的技术支持

如果您是使用任何学生版 IBM SPSS 软件产品的学生，请访问我们专为学生提供的在线教育解决方案 (http://www.ibm.com/spss/rd/students/)页面。如果您是使用大学提供的 IBM SPSS 软件副本的学生，请联系所在大学的 IBM SPSS 产品协调员。

客户服务

如果对发货或帐户存在任何问题，请联系您当地的办事处。请先准备好您的序列号以供识别。

培训讲座

IBM Corp. 提供公开的以及现场的培训讲座。所有讲座都是以实践小组为特色的。讲座将定期在各大城市开展。有关这些讲座的更多信息，请前往 http://www.ibm.com/software/analytics/spss/training。

附加出版物

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

内容

部分 I: 用户指南

部分 I:
用户指南

章 1

直销

“直销”选项提供了一组精心设计以改善直销活动效果的工具，它可以标识那些用于定义不同消费者群体的人口统计学、购买和其他特征，针对特定目标群体最大限度地提高正面响应率。

RFM 分析。 此方法标识那些最有可能对新产品做出响应的现有客户。 有关详细信息，请参阅第 2 页码第 2 章中的RFM 分析。

聚类分析。 这是一个用于揭示数据中的自然分组（或聚类）的探索性工具。例如，它可以根据各种人口统计和购买特征识别不同的客户组。 有关详细信息，请参阅第 13 页码第 3 章中的聚类分析。

潜在客户概要文件。 此方法使用先前或检验活动的结果来创建描述概要文件。您可以使用概要文件在未来的活动中集中面向特定的联系人群体。 有关详细信息，请参阅第 17 页码第 4 章中的潜在客户概要文件。

邮政编码响应率。 此方法使用先前活动的结果来计算邮政编码响应率。这些响应率可以用于在未来的活动中集中面向特定的邮政编码。 有关详细信息，请参阅第 23 页码第 5 章中的邮政编码响应率。

购买倾向。 此方法使用测试邮件或先前活动的结果来生成倾向得分。这些得分显示哪些联系人最有可能做出响应。 有关详细信息，请参阅第 30 页码第 6 章中的购买倾向。

控制包装检验。 此方法比较市场营销活动，以检查不同包装或商品之间是否存在显著的效果差异。 有关详细信息，请参阅第 37 页码第 7 章中的控制包装检验。

章 2

RFM 分析

RFM 分析是一种用于标识最可能对新产品做出反应的现有客户的方法。此方法常用于直销。RFM 分析基于以下简单理论：

- 标识最可能对新产品做出反应的现有客户的最重要因子是**崭新**。最近购买的客户比过去购买的客户更可能再次购买。
- 第二个重要的因子是**频率**。过去购买次数较多的客户比购买次数少的客户更可能做出反应。
- 第三个重要的因子是消费的总金额，称为**金额**。过去消费金额较多（所有购买的总和）的客户比消费金额较少的客户更可能做出反应。

RFM 分析的工作原理

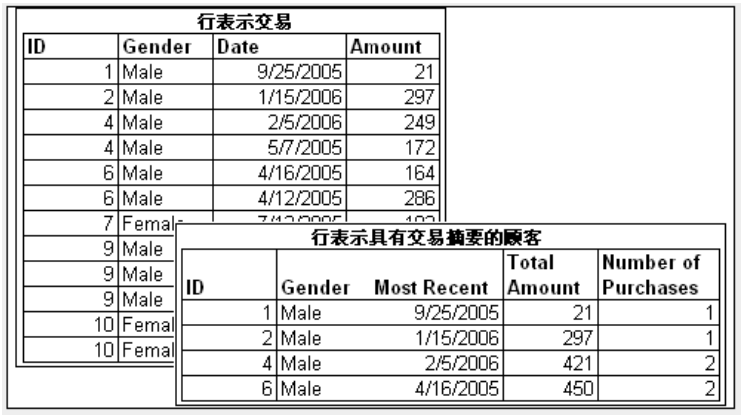
- 基于最近购买日期或自最近购买以来的时间间隔，为客户分配一个崭新得分。此得分基于将崭新值简单评级为少量类别。例如，如果您使用五个类别，则拥有最近购买日期的客户将获得崭新等级 5，而拥有过去购买日期的客户将获得崭新等级 1。
- 类似地，随后将为客户分配一个频率等级，其中较高的值代表购买频率较高。例如，在五个类别等级设计中，最常购买的客户将获得频率等级 5。
- 最后，按消费金额的值对客户进行评级，其中消费金额值最高的客户将获得最高等级。继续五个类别的示例，消费最多的客户将获得消费金额等级 5。

结果是每个客户获得四个得分：崭新、频率、金额以及合并 RFM 得分，即将三个单个得分连接为一个值。拥有最高合并 RFM 得分的客户即为“最佳”客户（最可能对产品做出反应的客户）。例如，在五个类别等级中，共有 125 种可能的合并 RFM 得分，最高合并 RFM 得分是 555。

数据注意事项

- 如果数据行代表交易（每行代表单笔交易，每个客户可能有多笔交易），则使用交易中的 RFM。 有关详细信息，请参阅第 3 页码来自交易数据的 RFM 得分。
- 如果数据行代表拥有所有交易摘要信息的客户（列包含消费的总金额、交易的总数和最近交易日期的值），则使用客户数据中的 RFM。 有关详细信息，请参阅第 4 页码来自客户数据的 RFM 得分。

图片 2-1
交易与客户数据



行表示交易

ID	Gender	Date	Amount
1	Male	9/25/2005	21
2	Male	1/15/2006	297
4	Male	2/5/2006	249
4	Male	5/7/2005	172
6	Male	4/16/2005	164
6	Male	4/12/2005	286
7	Femal		
9	Male		
9	Male		
9	Male		
10	Femal		
10	Femal		

行表示具有交易摘要的顾客

ID	Gender	Most Recent	Total Amount	Number of Purchases
1	Male	9/25/2005	21	1
2	Male	1/15/2006	297	1
4	Male	2/5/2006	421	2
6	Male	4/16/2005	450	2

来自交易数据的 RFM 得分

数据注意事项

数据集必须包含含有以下信息的变量：

- 标识每个个案（客户）的变量或变量组合。
- 拥有每次交易日期的变量。
- 拥有每次交易的消费金额值的变量。

图片 2-2
RFM 交易数据

ID	Date	Amount
1	08/04/2005	129
1	10/25/2004	50
1	07/24/2004	118
1	07/24/2004	136
1	09/04/2006	52
2	09/23/2005	183
2	11/05/2004	24
2	11/10/2005	66
2	12/03/2004	77
3	06/04/2005	102
3	05/15/2005	131

创建来自交易数据的 RFM 得分

▶ 从菜单中选择：
直销 > 选择方法

▶ 选择**帮助标识我的最佳联系人 (RFM 分析)**，并单击**继续**。

▶ 选择**交易数据**，然后单击**继续**。

图片 2-3
交易数据，“变量”选项卡

- 选择包含交易日期的变量。
- 选择包含每次交易的消费金额的变量。
- 选择汇总每个客户交易金额的方法：总数（所有交易总和）、均值、中位数或最大值（最高交易金额）。
- 选择唯一标识每个客户的变量或变量组合。例如，可以通过唯一 ID 代码或姓名组合来识别个案。

来自客户数据的 RFM 得分

数据注意事项

数据集必须包含含有以下信息的变量：

- 最近购买日期或自最近购买日期以来的时间间隔。这将用于计算崭新得分。
- 购买总次数。这将用于计算频率得分。
- 所有购买的摘要消费金额值。这将用于计算消费金额得分。通常，这是所有购买的总和（总数），但也可能是均值（平均值）、最大值（最大金额）或其他摘要测量。

图片 2-4
RFM 客户数据

ID	TotalAmount	MostRecent	NumberOfPurchases
1	485.00	09/04/2006	5
2	350.00	11/10/2005	4
3	233.00	06/04/2005	2
4	936.00	08/18/2006	7
5	359.00	07/07/2006	3
6	249.00	07/16/2006	3
7	1089.00	02/15/2006	7
8	423.00	08/21/2006	4
9	689.00	08/31/2006	7
10	325.00	10/13/2005	3

如果您想将 RFM 得分写入一个新的数据集，活动数据集还必须包含一个标识每个个案（客户）的变量或变量组合。

创建来自客户数据的 RFM 得分

- 从菜单中选择：
  直销 > 选择方法
- 选择**帮助标识我的最佳联系人（RFM 分析）**，并单击**继续**。
- 选择**客户数据**，然后单击**继续**。

图片 2-5
客户数据，“变量”选项卡

- 选择包含最近交易日期或代表自最近交易以来的时间间隔的数字的变量。

- 选择包含每个客户交易总次数的变量。
- 选择包含每个客户摘要消费金额的变量。
- 如果您想将 RFM 得分写入一个新的数据集，请选择唯一标识每个客户的变量或变量组合。例如，可以通过唯一 ID 代码或姓名组合来识别个案。

RFM 离散化

将大量数值分组为小量类别的过程有时也称为**离散化**。在 RFM 分析中，块是已评级的类别。您可以使用“离散化”选项卡修改用于将崭新、频率和消费金额值分配到这些块中的方法。

图片 2-6
“RFM 离散化”选项卡

离散化方法

嵌套。 在嵌套离散化中，简单等级被分配到崭新值。在每个崭新等级中，客户会分配到一个频率等级，然后在每个频率等级中，客户会分配到一个消费金额等级。这可以使合并 RFM 得分的分布更平均，但其缺点是会使频率和消费金额等级得分更难解释。例如，拥有崭新等级 5 的客户的频率等级 5 与拥有崭新等级 4 的客户的频率等级 5 意义是不同的，因为频率等级取决于崭新等级。

独立。 简单等级被分配到崭新、频率和消费金额值。三个等级独立分配。三个 RFM 组件中每个组件的解释因此都非常明确；一个客户的频率得分 5 与另一个客户的频率得分 5 意义是相同的，无论其崭新得分如何。对于较小的样本，这样做的缺点是会导致合并 RFM 得分的分布不平均。

块数

用于每个组件创建 RFM 得分的类别（块）数。可能的合并 RFM 得分的总数是三个值的乘积。例如，5 个崭新块、4 个频率块和 3 个消费金额块将创建总共 60 个可能的合并 RFM 得分，范围从 111 到 543。

- 每个组件的缺省块数是 5，将创建 125 个可能的合并 RFM 得分，范围从 111 到 555。
- 每个得分组件允许的最大块数是 9。

结

“同数”是两个或更多相等的崭新、频率或消费金额值。理想状况下，您希望在每个块中拥有大致相同的客户数量，但是大量同数的值可能影响块的分布。有两种方法可以处理同数：

- **将同数分配到相同的块**。无论对块分布的影响如何，此方法始终将同数的值分配到相同的块。这还提供了一致的离散化方法：如果两个客户具有相同的崭新值，那么他们将始终分配到相同的崭新得分。但是在极端示例中，您可能有 1,000 个客户，其中 500 个在同一天进行了最近一次购买。在 5 块评级中，50% 的客户因此获得了崭新得分 5，而非所需的 20%。

  注意，使用嵌套离散化方法时，“一致性”对于频率和消费金额得分有点过于复杂了，因为要在崭新得分块中分配频率得分，在频率得分块中分配消费金额得分。因此，无论同数的值如何处理，拥有相同频率值的两个客户如果没有相同的崭新得分，他们仍无法获得相同的频率得分。

- **随机分配同数。** 此操作通过在评级前将非常小的随机方差因子分配给同数来确保块的平均分布；因此为了将值分配给已评级的块，不存在同数的值。此过程对原始值没有影响。只用于消除同数。尽管这可以使块分布平均（每个块中的客户数大致相同），但对于似乎拥有类似或相同崭新、频率和/或消费金额值的客户仍可能导致完全不同的得分结果。这在客户总数相对较少和/或同数的数量相对较多时尤为明显。

表 2-1
将同数分配给相同的块以及随机分配同数

ID	最近购买（崭新）	将同数分配给相同的块	随机分配同数
1	10/29/2006	5	5
2	10/28/2006	4	4
3	10/28/2006	4	4
4	10/28/2006	4	5
5	10/28/2006	4	3
6	9/21/2006	3	3
7	9/21/2006	3	2
8	8/13/2006	2	2
9	8/13/2006	2	1
10	6/20/2006	1	1

- 在本例中，将同数分配给相同的块导致块分布不均：5 (10%)，4 (40%)，3 (20%)，2 (20%)，1 (10%)。
- 随机分配同数导致每个块中分到 20%，但是要获得此结果，需将拥有日期值 10/28/2006 的四个个案分配给三个不同的块，并将拥有日期值 8/13/2006 的两个个案也分配给不同的块。

  注意，将同数分配给不同的块的方式完全随机（受结果要和每个块中的个案数相等的约束）。如果您用相同方法计算第二个得分集合，则拥有同数的值的任何特定个案的等级可能更改。例如，个案 4 的崭新等级 5 和个案 5 的崭新等级 3 可能被第二次切换。

保存来自交易数据的 RFM 得分

来自交易数据的 RFM 始终以每个客户一行的方式创建新汇总数据集。使用“保存”选项卡来指定您想保存的得分和其他变量以及它们的保存位置。

图片 2-7
交易数据，“保存”选项卡

变量

唯一标识每个客户的 ID 变量被自动保存在新数据集中。在新数据集中可以保存以下附加变量：

- **每个客户最近交易的日期。**
- **交易次数。** 每个客户交易行的总数。
- **金额。** 每个客户的摘要金额（基于您在“变量”选项卡上选择的摘要方法）。
- **崭新得分。** 分配给每个客户的基于最近交易日期的得分。得分越高表示交易日期越近。

- **频率得分。** 分配给每个客户的基于交易总数的得分。得分越高表示交易越多。
- **消费金额得分。** 分配给每个客户的基于所选消费金额摘要测量的得分。得分越高表示消费金额摘要测量的值越高。
- **RFM 得分。** 三个单个得分合为一个值：(崭新得分 x 100) + (频率得分 x 10) + 消费金额得分。

缺省情况下，所有可用变量都包括在新数据集中；因此，取消选择（取消选中）您不想包括的变量。根据需要，您可以指定自己的变量名称。变量名称必须符合标准变量命名规则。

位置

来自交易数据的 RFM 始终以每个客户一行的方式创建新汇总数据集。您可以在当前会话中创建新数据集或在外部数据文件中保存 RFM 得分数据。数据集名称必须符合标准变量命名规则。（此限制不适用于外部数据文件名称。）

保存来自客户数据的 RFM 得分

对于客户数据，您可以将 RFM 得分变量添加到活动数据集或创建一个包含选定得分变量的新数据集。使用“保存”选项卡来指定您想保存的得分变量及其保存位置。

图片 2-8
客户数据，“保存”选项卡

保存的变量名称

- **自动生成唯一的名称。** 这样可确保在将得分变量添加到活动数据集时，新变量名称为唯一。如果您想将多个不同的 RFM 得分集合（基于不同标准）添加到活动数据集，这一点尤其有用。
- **自定义名称。** 这允许您将自己的变量名称分配到得分变量。变量名称必须符合标准变量命名规则。

变量

选择（选中）想要保存的得分变量：

- **崭新得分。** 分配给每个客户的基于在“变量”选项卡上选定的“交易日期”或“间隔”变量的值的得分。日期越近或间隔值越低分配到的得分越高。
- **频率得分。** 分配给每个客户的基于在“变量”选项卡上选定的“交易数”变量的得分。值越高分配到的得分越高。
- **消费金额得分。** 分配给每个客户的基于在“变量”选项卡上选定的“金额”变量的得分。值越高分配到的得分越高。
- **RFM 得分。** 三个单个得分合为一个值：(崭新得分*100)+(频率得分*10)+消费金额得分。

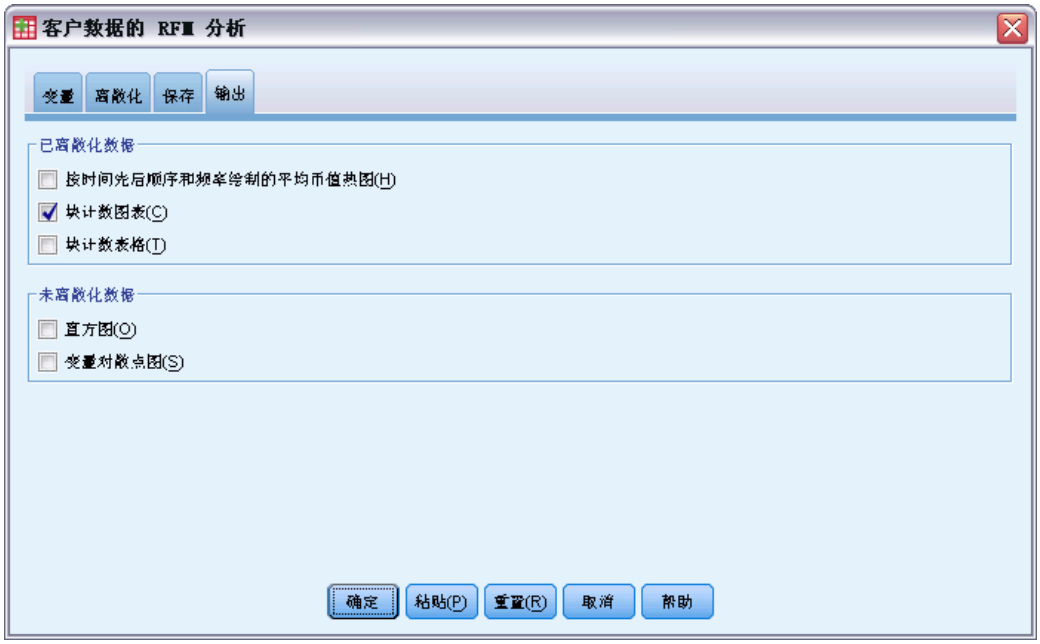
位置

对于客户数据，您有三个位置可以保存新的 RFM 得分：

- **活动数据集。** 将选定 RFM 得分变量添加到活动数据集。
- **新数据集。** 选定的 RFM 得分变量和唯一标识每个客户（个案）的 ID 变量将被写入到当前会话中的新数据集。数据集名称必须符合标准变量命名规则。 仅当您在“变量”选项卡上选择了一个或多个“客户标识”变量时，此选项才可用。
- **文件。** 选定的 RFM 得分和唯一标识每个客户（个案）的 ID 变量将保存在外部数据文件中。仅当您在“变量”选项卡上选择了一个或多个“客户标识”变量时，此选项才可用。

RFM 输出

图片 2-9
“RFM 输出”选项卡



已离散化数据

离散化数据的图表基于计算的崭新、频率和消费金额得分。

按崭新和频率绘制的消费金额均值热图。 消费金额均值分布热图显示由崭新和频率得分定义的类别的消费金额均值。颜色越深的区域表示消费金额均值越高。

块计数图。 块计数图表显示选定离散化方法的块分布。每个条代表将被分配每个合并 RFM 得分的个案数。

- 尽管您通常希望相当均匀的分布，即所有（或多数）条大体高度相同，但当使用将同数的值分配给相同块的默认离散化方法时，必然会产生一定量的偏差。
- 块分布中的极值波动和/或较多空的块可能表明您应尝试另一种离散化方法（块数量和/或随机分配结数量较少），或重新考虑 RFM 分析的适用性。

块计数表。 与块计数图中的信息相同，不同之处在于以表格形式呈现，每个单元格中为块计数。

未离散化数据

未离散化数据的图表基于用来创建崭新、频率和消费金额得分的原始变量。

直方图。 直方图显示用于计算崭新、频率和消费金额得分的三个变量的值的相对分布。这些直方图经常用来表示正态或对称分布以外的偏斜分布。

每个直方图的水平轴始终采用左侧为较小值、右侧为较大值的顺序。但对于崭新，图表的解释依赖于崭新测量的类型：日期或时间间隔。对于日期，左侧条代表更“旧”的值（即较远日期比较近日期的值更小）。对于时间间隔，左侧条代表更“新”的值（即时间间隔越小，交易离现在越近）。

变量对散点图。 这些散点图显示用于计算崭新、频率和消费金额得分的三个变量之间的关系。

常常会看到频率刻度上明显的多点线性分组，因为频率常常代表相对小范围的离散值。例如，如果交易总数不超过 15，则只有 15 个可能的频率值（除非你计入了不算一次的零散交易），尽管可能有数百个可能的崭新值和数千个消费金额值。

崭新轴的解释依赖于崭新测量的类型：日期或时间间隔。对于日期，越接近原点的点代表离现在越远的过去日期。对于时间间隔，越接近原点的点代表越“新”的值。

章 3

聚类分析

聚类分析是用于揭示数据中的自然分组（或聚类）的探索性工具。例如，它可以根据各种人口统计和购买特征识别不同的客户组。

示例。零售和消费者产品公司定期地对描述客户的购买习惯、性别、年龄、收入水平等的数据应用聚类技术。这些公司为每个消费者群体设计营销和产品开发战略，以增加销售额和建立品牌忠诚度。

聚类分析数据注意事项

数据。 此过程既处理连续字段也处理分类字段。每个记录（行）代表要聚类的客户，字段（变量）代表聚类所基于的属性。

记录顺序。 注意，结果可取决于记录顺序。为使顺序的影响降至最低，您可能会考虑随机排序记录。您可能想通过以不同随机顺序排序的记录来多次运行分析，以验证给定解的稳定性。

测量级别。 正确指定测量级别是非常重要的，因为它会影响结果计算。

- **标定.** 当变量值表示不具有内在等级的类别时，该变量可以作为名义变量；例如，雇员任职的公司部门。名义变量的示例包括地区、邮政编码和宗教信仰。
- **有序.** 当变量值表示带有某种内在等级的类别时，该变量可以作为有序变量；例如，从十分不满意到十分满意的服务满意度水平。有序变量的示例包括表示满意度或可信度的态度分数和优先选择评分。
- **连续.** 当变量值表示带有有意义的度规的已排序类别时，该变量可以作为刻度（连续）变量对待，以便在值之间进行合适的距离比较。刻度变量的示例包括以年为单位的年龄和以千美元为单位的收入。

每个字段旁的图标指示当前的测量级别。

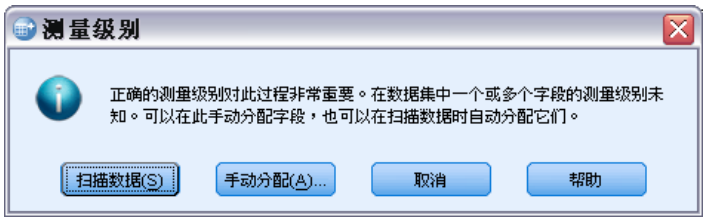
	数值	字符串	日期	时间
尺度（连续）		n/a		
有序				
名义				

可以在数据编辑器的“变量视图”中更改测量级别，或者也可以使用“定义变量属性”对话框为每个字段建议适当的测量级别。

具有未知测量级别的字段

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

图片 3-1
测量级别警报



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

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

获取聚类分析

从菜单中选择：
直销 > 选择方法

▶ 选择**将我的联系人分段到聚类**。

图片 3-2
“聚类分析字段”选项卡

- 选择要用于创建段的分类（名义、有序）字段和连续（尺度）字段。
- 单击运行以运行该过程。

设置

图片 3-3
“聚类分析设置”选项卡

“设置”选项卡允许您显示或不显示描述段的图表和表格，在数据集中保存新字段以标识数据集中每个记录的段（聚类），以及指定在聚类解中包含多少个段。

显示图表和表格。 显示描述段的表格和图表。

段成员。 保存新字段（变量），以标识每个记录所属的段。

- 字段名必须符合 IBM® SPSS® Statistics 命名规则。
- 段关系字段名不能与数据集中现有字段名重复。如果在同一数据集上多次运行此过程，则需要每次指定不同的名称。
- **段的数量。** 控制如何确定段的数量。
- **自动确定。** 该过程将自动确定“最佳”的段数量，但应低于指定的最大数量。

指定固定值。 该过程将生成指定数量的段。

章 4

潜在客户概要文件

此方法使用先前或检验活动的结果来创建描述概要文件。您可以使用概要文件在未来的活动中集中面向特定的联系人群体。响应字段显示谁对先前或检验活动做出了响应。概要文件列表包含您打算用来创建概要文件的特征。

示例。 根据测试邮件的结果，公司直销部门想要生成以人口统计信息为基础的最可能对产品做出响应的客户类型概要文件。

输出

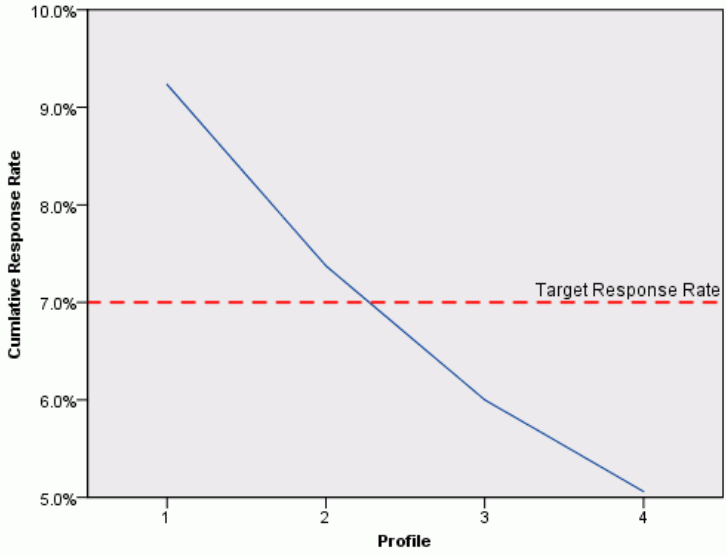
输出包括一个表格，提供每个概要文件组的说明，并显示响应率（正响应的百分比）和累积响应率，此外还包括一个累积响应率图表。如果包含目标最低响应率，则表格将通过颜色编码显示哪些概要文件满足最低累积响应率，同时图表将在指定的最低响应率值处包含参考线。

图片 4-1
响应率表格和图表

响应率

数字	概要文件			
	描述	组大小	响应率	累积响应率
1	Region = "West","South","East" Gender = "Female" Married = "No"	379	9.2%	9.2%
2	Region = "West","South","East" Gender = "Female" Married = "Yes"	299	5.0%	7.4%
3	Region = "West","South","East" Gender = "Male"	722	4.7%	6.0%
4	Region = "North"	517	2.5%	5.1%

绿色: 满足目标响应率。
红色: 不满足目标响应率。



潜在客户概要文件数据注意事项

响应字段。 响应字段必须为名义或有序字段。它可以是字符串或数值。如果此字段包含有指示购买数量或金额的值，您将需要创建新的字段，并使其中一个值代表所有正响应。有关详细信息，请参阅第 21 页码创建分类响应字段。

正响应值。正响应值标识那些做出正面响应的客户（例如，购买产品）。所有其他非缺失响应值均被假设为表示负响应。如果为响应字段定义有值标签，则会在下拉列表中显示这些值标签。

创建概要文件。 这些字段可以为名义、有序或连续（尺度）字段。它们可以是字符串或数值。

测量级别。 正确指定测量级别是非常重要的，因为它会影响结果计算。

- **标定**. 当变量值表示不具有内在等级的类别时，该变量可以作为名义变量；例如，雇员任职的公司部门。名义变量的示例包括地区、邮政编码和宗教信仰。
- **有序**. 当变量值表示带有某种内在等级的类别时，该变量可以作为有序变量；例如，从十分不满意到十分满意的服务满意度水平。有序变量的示例包括表示满意度或可信度的态度分数和优先选择评分。
- **连续**. 当变量值表示带有有意义的度规的已排序类别时，该变量可以作为刻度（连续）变量对待，以便在值之间进行合适的距离比较。刻度变量的示例包括以年为单位的年龄和以千美元为单位的收入。

每个字段旁的图标指示当前的测量级别。

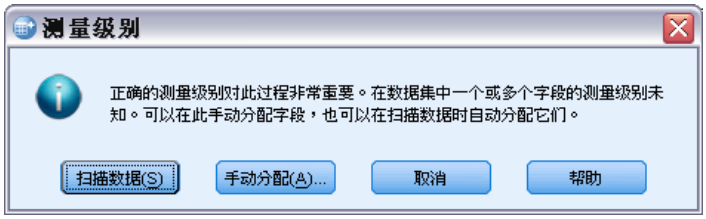
	数值	字符串	日期	时间
尺度（连续）		n/a		
有序				
名义				

可以在数据编辑器的“变量视图”中更改测量级别，或者也可以使用“定义变量属性”对话框为每个字段建议适当的测量级别。

具有未知测量级别的字段

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

图片 4-2
测量级别警报



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

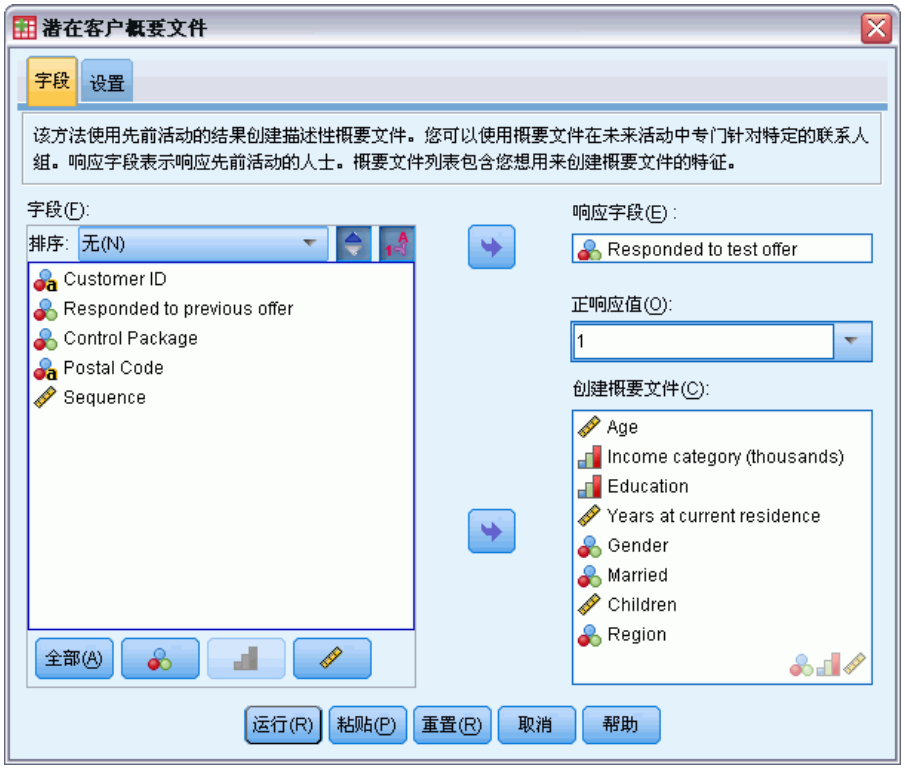
由于测量级别对该过程很重要，因此您无法访问运行该过程的对话框，除非所有字段均定义了测量级别。

获取潜在客户概要文件

从菜单中选择：
直销 > 选择方法

- 选择生成对产品做出响应的我的联系人的概要文件。

图片 4-3
“潜在客户概要文件：字段”选项卡



- 选择标识哪些联系人对产品做出响应的字段。该字段必须为名义或有序字段。
- 输入表示正响应的值。如果有任何值定义了值标签，则可以从下拉列表中选择值标签，这将显示对应的值。
- 选择用于创建概要文件的字段。
- 单击运行以运行该过程。

设置

图片 4-4
“潜在客户概要文件：设置”选项卡

“设置”选项卡允许您控制最小概要文件组大小，以及在输出中包含最低响应率阈值。

最小概要文件组大小。 每个概要文件代表数据集中一组联系人的公共特征（例如，居住在西部、年龄在 40 岁以下的女性）。默认情况下，最小概要文件组大小为 100。组大小越小，可以显示的组越多；但组大小越大，结果会越可靠。值必须为正整数。

在结果中包括最小响应率阈值信息。 结果包括一个表格，显示响应率（正响应的百分比）和累积响应率，此外还包括一个累积响应率图表。如果输入目标最低响应率，则表格将通过颜色编码显示哪些概要文件满足最低累积响应率，同时图表将在指定的最低响应率值处包含参考线。 值必须大于 0 且小于 100。

创建分类响应字段

响应字段应当为分类字段，且具有一个代表所有正响应的值。任何其他非缺失值均被假设为负响应。如果响应字段代表连续（尺度）值，例如购买数量或购买金额，则需要创建新字段，以便为所有非零响应值指定单个正响应值。

- 如果负响应被记录为 0（不是空值，后者作为缺失处理），则可以通过以下公式进行计算：

  ```
  NewName=OldName>0
  ```

其中 NewName 为新字段的名称，OldName 为原始字段的名称。这是一个逻辑表达式，它为所有大于零的非缺失值指定值 1，为所有小于或等于零的非缺失值指定值 0。

- 如果未记录有负响应值，这些值将作为缺失处理，公式则更为复杂：

  NewName=NOT(MISSING(OldName))

  在此逻辑表达式中，为所有非缺失响应值指定值 1，为所有缺失响应值指定值 0。

- 如果不能区分负 (0) 响应值和缺失值，则无法计算准确的响应值。如果实际缺失值相对较少，这对计算的响应率可能并无显著影响。但如果缺失值较多，比如当仅为整个数据集中少量检验样本记录响应信息时，则计算的响应率将没有意义，因为它们将明显低于实际响应率。

创建分类响应字段

▶ 从菜单中选择：
转换 > 计算变量

▶ 为“目标变量”输入新的字段（变量）名称。

▶ 如果负响应被记录为 0，则为“数值表达式”输入 **OldName>0**，其中 OldName 为原始字段名。

▶ 如果负响应被记录为缺失（空），则为“数值表达式”输入 **NOT(MISSING(OldName))**，其中 OldName 为原始字段名。

章 5

邮政编码响应率

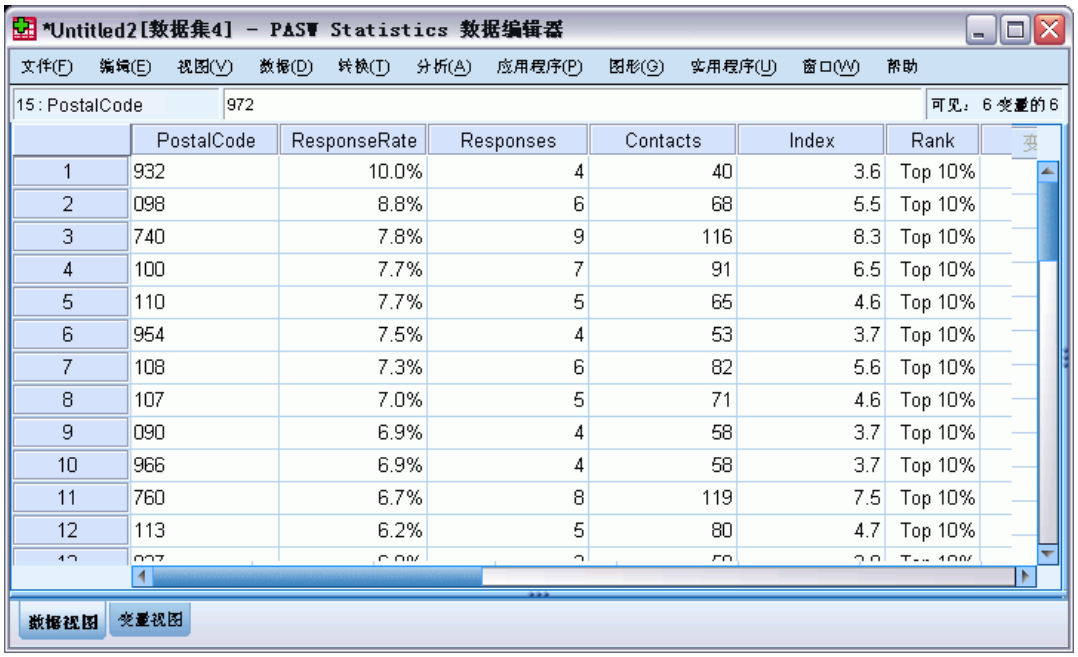
此方法使用先前活动的结果来计算邮政编码响应率。这些响应率可以用于在未来的活动中集中面向特定的邮政编码。响应字段显示谁对先前活动做出了响应。邮政编码字段标识包含邮政编码的字段。

示例。 根据先前邮件的结果，公司直销部门按邮政编码生成响应率。然后，根据不同的标准，例如最低可接受响应率和/或在邮件中包括的最大联系人数量，他们可以集中面向特定的邮政编码。

输出

此过程的输出包含一个内容为邮政编码响应率的新数据集，以及按十分位数排序（前 10%，前 20%，等等）列出结果摘要的表格和图表。表格可以基于用户指定的最低累积响应率或最大联系人数量进行颜色编码。

图片 5-1
包括邮政编码响应率的数据集



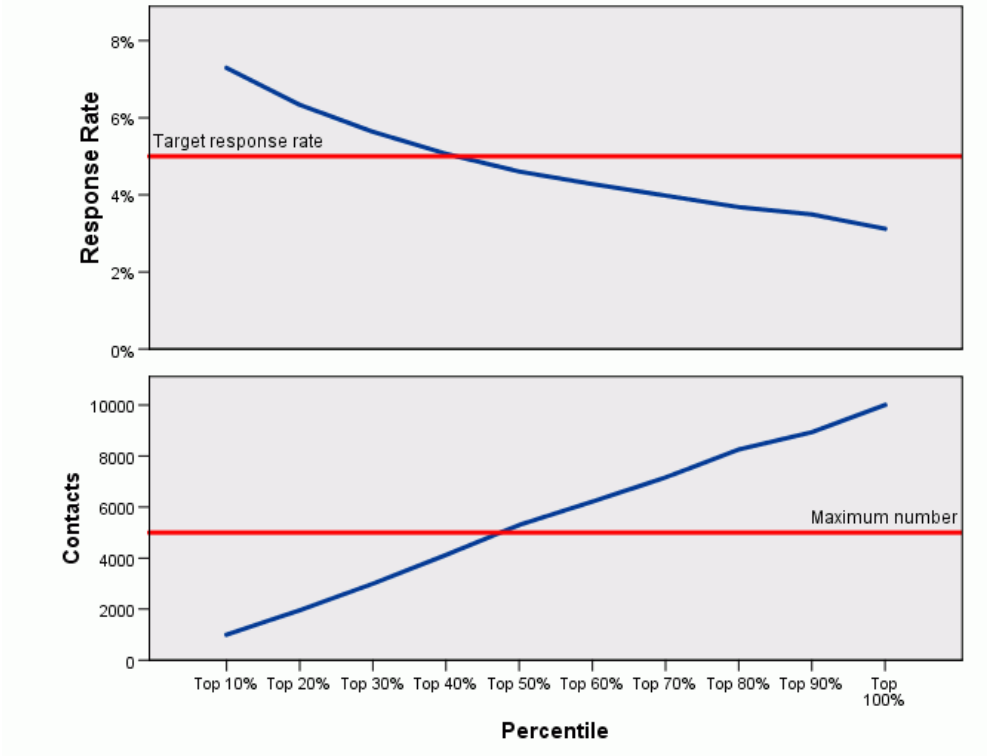
	PostalCode	ResponseRate	Responses	Contacts	Index	Rank
1	932	10.0%	4	40	3.6	Top 10%
2	098	8.8%	6	68	5.5	Top 10%
3	740	7.8%	9	116	8.3	Top 10%
4	100	7.7%	7	91	6.5	Top 10%
5	110	7.7%	5	65	4.6	Top 10%
6	954	7.5%	4	53	3.7	Top 10%
7	108	7.3%	6	82	5.6	Top 10%
8	107	7.0%	5	71	4.6	Top 10%
9	090	6.9%	4	58	3.7	Top 10%
10	966	6.9%	4	58	3.7	Top 10%
11	760	6.7%	8	119	7.5	Top 10%
12	113	6.2%	5	80	4.7	Top 10%

图片 5-2
摘要表格和图表

响应率

Percentile	响应率	Contacts	Cumulative Response Rate	Total Contacts
Top 10%	7.3	1001	7.3	1001
Top 20%	5.3	956	6.3	1957
Top 30%	4.3	1042	5.6	2999
Top 40%	3.5	1127	5.1	4126
Top 50%	3.0	1173	4.6	5299
Top 60%	2.4	914	4.3	6213
Top 70%	2.0	948	4.0	7161
Top 80%	1.7	1095	3.7	8256
Top 90%	1.2	680	3.5	8936
Top 100%	.0	1064	3.1	10000

Green Red Caption



新数据集包含以下字段：

- **邮政编码。** 如果邮政编码组仅基于完整值的某个部分，则为该部分邮政编码的值。在 Excel 文件中，此列的标题行标签为原始数据集中的邮政编码字段名称。
- **响应率。** 每个邮政编码中正响应的百分比。
- **响应。** 每个邮政编码中正响应的个数。

- **联系人。** 在每个邮政编码中包含响应字段的非缺失值的联系人总数。
- **指标。**基于公式 N x P x (1-P) 的“加权”响应，其中 N 为联系人数量，P 为以比例表示的响应率。
- **排序。** 以降序排列的累积邮政编码响应率的十分位数排序（前 10%，前 20%，等等）。

邮政编码响应率数据注意事项

响应字段。 响应字段可以是字符串或数值。如果此字段包含有指示购买数量或金额的值，您将需要创建新的字段，并使其中一个值代表所有正响应。有关详细信息，请参阅第 28 页码创建分类响应字段。

正响应值。正响应值标识那些做出正面响应的客户（例如，购买产品）。所有其他非缺失响应值均被假设为表示负响应。如果为响应字段定义有值标签，则会在下拉列表中显示这些值标签。

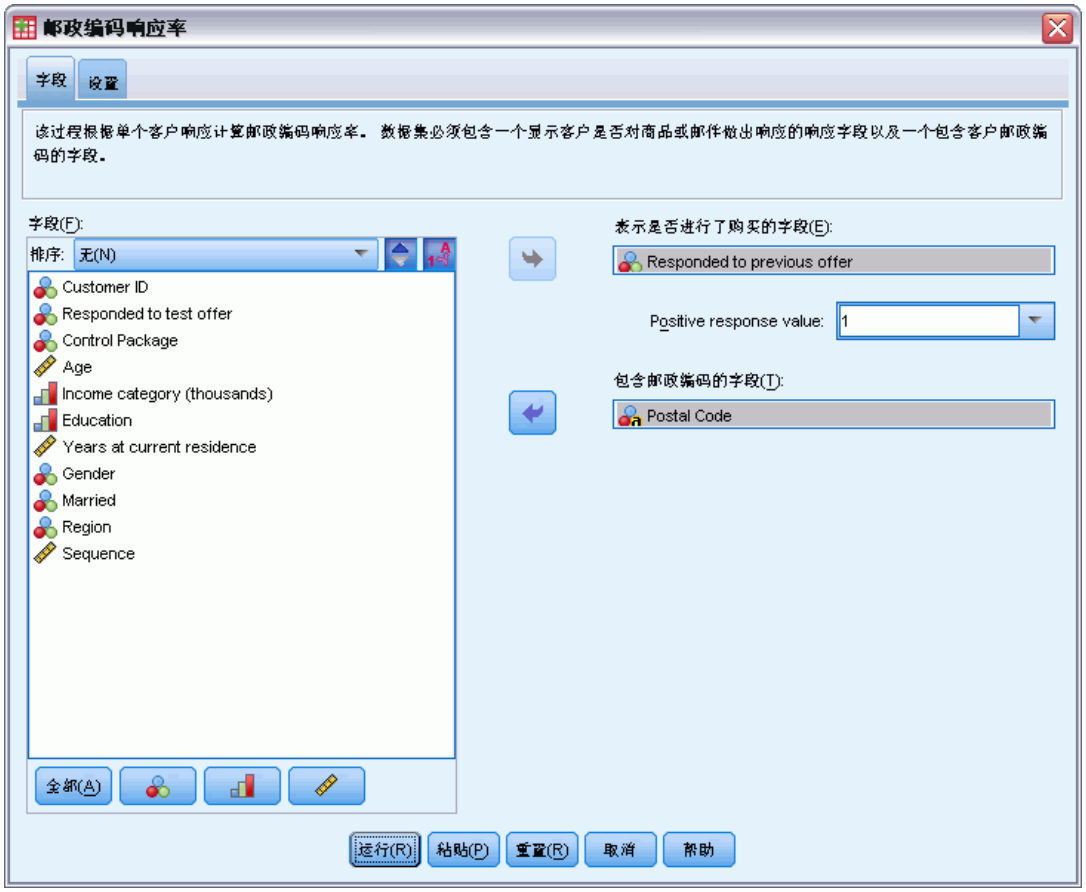
邮政编码字段。 邮政编码字段可以是字符串或数值。

获取邮政编码响应率

从菜单中选择：
直销 > 选择方法

▶ 选择**标识最佳响应邮政编码**。

图片 5-3
“邮政编码响应率：字段”选项卡



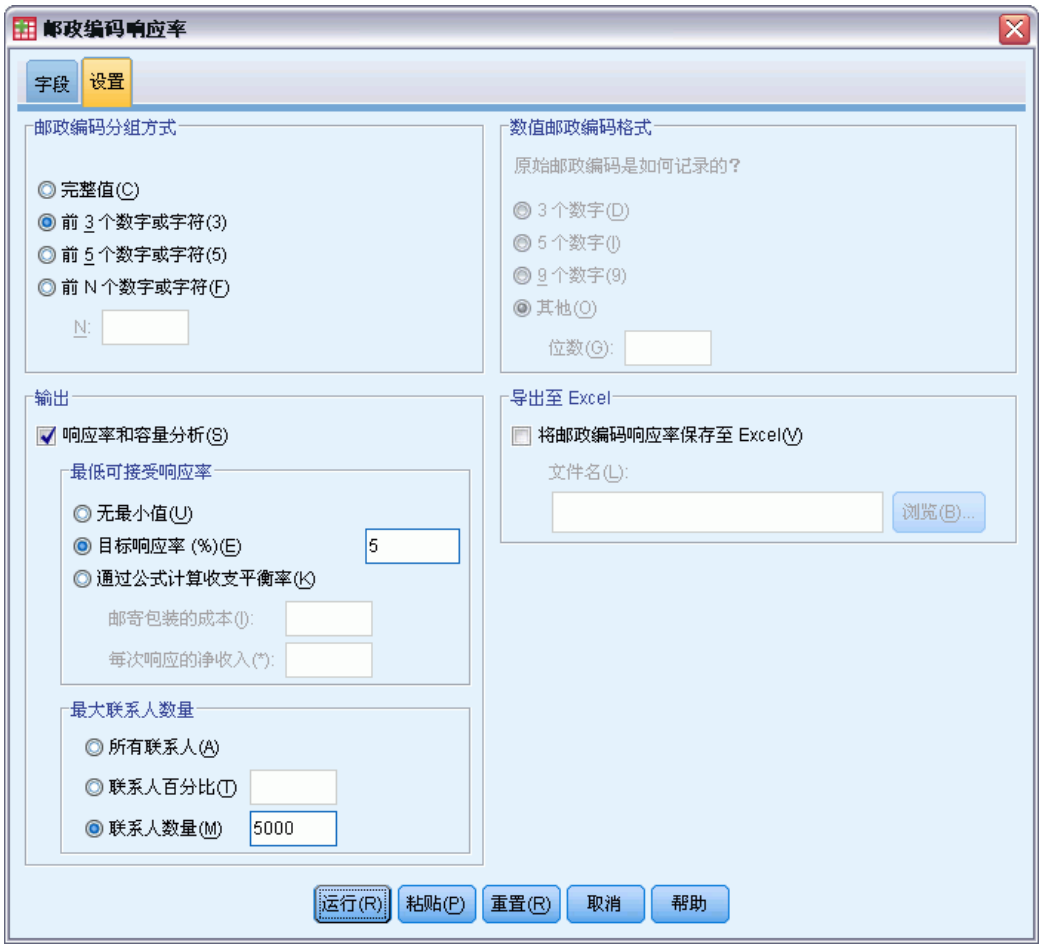
- 选择标识哪些联系人对产品做出响应的字段。
- 输入表示正响应的值。如果有任何值定义了值标签，则可以从下拉列表中选择值标签，这将显示对应的值。
- 选择包含邮政编码的字段。
- 单击**运行**以运行该过程。

根据需要，您可以：

- 根据邮政编码的前 n 个字符或数字而不是整个值生成响应率
- 自动将结果保存到 Excel 文件
- 控制输出显示选项

设置

图片 5-4
“邮政编码响应率：设置”选项卡



邮政编码分组方式

这可以确定如何分组记录以计算响应率。缺省情况下，使用整个邮政编码，并将具有相同邮政编码的所有记录分组在一起以计算组响应率。或者，也可以基于完整邮政编码的某个部分，即前 n 个数字或字符来对记录分组。例如，您可能打算基于 10 位字符邮政编码的前 5 位字符，或 5 位数字邮政编码的前 3 位数字来对记录分组。输出数据集将为每个邮政编码组包含一条记录。如果输入值，则此值必须为正整数。

数值邮政编码格式

如果邮政编码字段为数值，并且您打算基于前 n 位数字而不是整个值来分组邮政编码，则需要指定原始值中的位数。这里的位数是邮政编码中的最大可能位数。例如，如果邮政编码字段包含 5 位和 9 位数字邮政编码的组合，则需要指定 9 作为位数。

注意：根据显示格式，5 位邮政编码可能在显示时只包含 4 位数字，这是因为前面的零被隐含。

输出

除了包含邮政编码响应率的新数据集外，还可以显示按十分位数排序（前 10%，前 20%，等等）列出结果摘要的表格和图表。表格显示每个十分位数分级中的响应率、累积响应率、记录数和累积记录数。图表则显示每个十分位数分级中的累积响应率和累积记录数。

最低可接受响应率。 如果输入目标最低响应率或收支平衡公式，则表格将通过颜色编码显示哪些十分位数分级满足最低累积响应率，同时图表将在指定的最低响应率值处包含参考线。

- **目标响应率。** 响应率表示为百分比（每个邮政编码组中正响应的百分比）。值必须大于 0 且小于 100。
- **通过公式计算收支平衡率。** 基于下列公式计算最低累积响应率：(邮寄包装的成本/每次响应的净收入) x 100。这两个值必须为正数。结果必须为大于 0 且小于 100 的值。例如，如果邮寄包装的成本为 $0.75，而每次响应的净收入为 $56，则最低响应率为：(0.75/56) x 100 = 1.34%。

最大联系人数量。 如果指定了最大联系人数量，则表格将通过颜色编码显示哪些十分位数分级未超过累积最大联系人数量（记录数），图表将在该值处包含参考线。

- **联系人百分比。** 以百分比表示的最大联系人数量。例如，您可能想知道具有最高响应率且包含不超过全部联系人的 50% 的十分位数分级。值必须大于 0 且小于 100。
- **联系人数量。** 以联系人数量表示的最大联系人数量。例如，如果您不愿邮寄超过 10,000 个包装，可以将值设置在 10000。此值必须为正整数（无分组符号）。

如果同时指定了最低可接受响应率和最大联系人数量，则表格颜色编码将基于最先满足的条件。

导出至 Excel

该过程自动创建一个包括邮政编码响应率的新数据集数据集中的每条记录（行）代表一个邮政编码。您可以自动将同一信息保存到 Excel 文件。此文件以 Excel 97-2003 格式保存。

创建分类响应字段

响应字段应当为分类字段，且具有一个代表所有正响应的值。任何其他非缺失值均被假设为负响应。如果响应字段代表连续（尺度）值，例如购买数量或购买金额，则需要创建新字段，以便为所有非零响应值指定单个正响应值。

- 如果负响应被记录为 0（不是空值，后者作为缺失处理），则可以通过以下公式进行计算：

  ```
  NewName=OldName>0
  ```

其中 NewName 为新字段的名称，OldName 为原始字段的名称。这是一个逻辑表达式，它为所有大于零的非缺失值指定值 1，为所有小于或等于零的非缺失值指定值 0。

- 如果未记录有负响应值，这些值将作为缺失处理，公式则更为复杂：

  NewName=NOT(MISSING(OldName))

  在此逻辑表达式中，为所有非缺失响应值指定值 1，为所有缺失响应值指定值 0。

- 如果不能区分负 (0) 响应值和缺失值，则无法计算准确的响应值。如果实际缺失值相对较少，这对计算的响应率可能并无显著影响。但如果缺失值较多，比如当仅为整个数据集中少量检验样本记录响应信息时，则计算的响应率将没有意义，因为它们将明显低于实际响应率。

创建分类响应字段

▶ 从菜单中选择：
转换 > 计算变量

▶ 为“目标变量”输入新的字段（变量）名称。

▶ 如果负响应被记录为 0，则为“数值表达式”输入 OldName>0，其中 OldName 为原始字段名。

▶ 如果负响应被记录为缺失（空），则为“数值表达式”输入 NOT(MISSING(OldName))，其中 OldName 为原始字段名。

章 6

购买倾向

购买倾向使用测试邮件或先前活动的结果来生成得分。这些得分显示哪些联系人最有可能做出响应。响应字段显示谁对测试邮件或先前活动做出了回应。倾向字段是您要用于预测具有类似特征的联系人将做出反应的可能性的特征。

此方法采用二元 Logistic 回归构建预测模型。构建并应用预测模型的过程包含两个基本步骤：

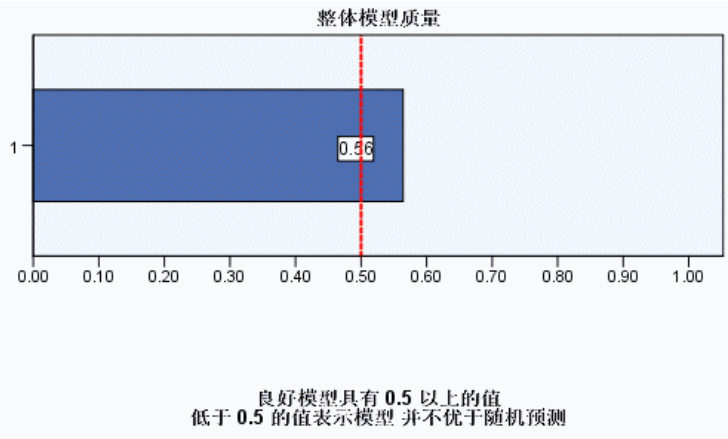
- 构建模型并保存模型文件。使用数据集构建兴趣结果（通常被称为**目标**）已知的模型。例如，如果您希望构建可预测谁可能会响应直接邮寄活动的模型，则需要从已包含响应人和未响应人信息的数据集开始。例如，这可能是对一小组客户发送的测试邮件的结果或来自过去类似活动的响应信息。
- 应用该模型到其他数据集（其中兴趣结果未知）以获取预测结果。

示例。 公司直销部门使用测试邮件的结果，为其联系人数据库的其余部分指定倾向得分，他们使用各种人口统计学特征来标识最有可能做出响应和购买产品的联系人。

输出

此过程自动在数据集中创建包含检验数据的倾向得分的新字段，以及可用于对其他数据集评分的 XML 模型文件。可选的诊断输出包括一个整体模型质量图表和一个比较预测响应与实际响应的分类表。

图片 6-1
整体模型质量图表



购买倾向数据注意事项

响应字段。 响应字段可以是字符串或数值。如果此字段包含有指示购买数量或金额的值，您将需要创建新的字段，并使其中一个值代表所有正响应。有关详细信息，请参阅第 35 页码创建分类响应字段。

正响应值。正响应值标识那些做出正面响应的客户（例如，购买产品）。所有其他非缺失响应值均被假设为表示负响应。如果为响应字段定义有值标签，则会在下拉列表中显示这些值标签。

预测倾向。 这些用于预测倾向的字段可以是字符串或数值。它们可以为名义、有序或连续（尺度）字段，但必须为所有预测变量字段指定适当的测量级别。

测量级别。 正确指定测量级别是非常重要的，因为它会影响结果计算。

- **标定.** 当变量值表示不具有内在等级的类别时，该变量可以作为名义变量；例如，雇员任职的公司部门。名义变量的示例包括地区、邮政编码和宗教信仰。
- **有序.** 当变量值表示带有某种内在等级的类别时，该变量可以作为有序变量；例如，从十分不满意到十分满意的服务满意度水平。有序变量的示例包括表示满意度或可信度的态度分数和优先选择评分。
- **连续.** 当变量值表示带有有意义的度规的已排序类别时，该变量可以作为刻度（连续）变量对待，以便在值之间进行合适的距离比较。刻度变量的示例包括以年为单位的年龄和以千美元为单位的收入。

每个字段旁的图标指示当前的测量级别。

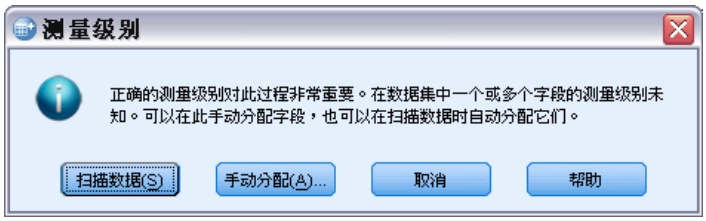
	数值	字符串	日期	时间
尺度（连续）		n/a		
有序				
名义				

可以在数据编辑器的“变量视图”中更改测量级别，或者也可以使用“定义变量属性”对话框为每个字段建议适当的测量级别。

具有未知测量级别的字段

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

图片 6-2
测量级别警报



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

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

获取购买倾向得分

从菜单中选择：
直销 > 选择方法

▶ 选择**选择最有可能购买的联系人**。

图片 6-3
“购买倾向：字段”选项卡

- 选择标识哪些联系人对产品做出响应的字段。
- 输入表示正响应的值。如果有任何值定义了值标签，则可以从下拉列表中选择值标签，这将显示对应的值。
- 选择用于预测倾向的字段。

保存模型 XML 文件以对其他数据文件评分：

- 选择（勾选）**将模型信息导出到 XML 文件**。
- 输入目录路径和文件名，或单击**浏览**导航到要用于保存模型 XML 文件的位置。
- 单击**运行**以运行该过程。

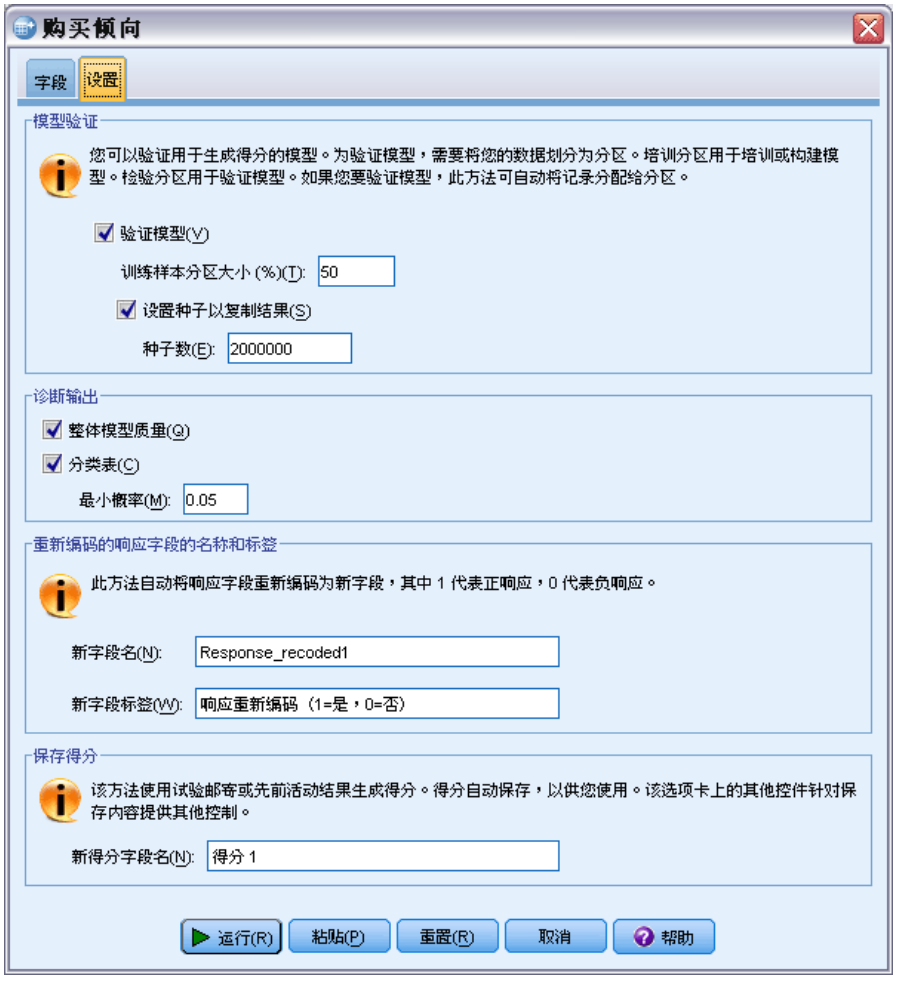
使用模型文件对其他数据集评分：

- 打开您要评分的数据集。

▶ 使用评分向导将模型应用到数据集。从菜单中选择：
实用程序 > 评分向导。

设置

图片 6-4
“购买倾向：设置”选项卡



模型验证

模型验证创建训练和测试组以供诊断用途。如果在“诊断输出”部分选择分类表，则此表将分为训练（选中）和测试（未选中）部分，以方便比较。仅当选择分类表后，才能选择模型验证。得分基于从训练样本生成的模型，而训练样本包含的记录数始终少于可用记录总数。例如，默认训练样本大小为 50%，在一半可用记录上构建的模型不如基于全部可用记录的模型可靠。

- **训练样本分区大小（%）。** 指定分配给训练样本的记录百分比。响应字段的具有非缺失值的其余记录则被指定给测试样本。值必须大于 0 且小于 100。
- **设置种子以复制结果。** 由于记录是随机分配给训练和测试样本的，因此每次运行过程时可能会得到不同的结果，除非始终指定相同的随机数种子值。

诊断输出

整体模型质量。 显示整体模型质量（表示为 0 和 1 之间的值）条形图。良好的模型应具有大于 0.5 的值。

分类表。 显示一个表格，对预测的正、负响应和实际的正、负响应进行比较。总体准确率可以在某些方面显示模型工作情况，不过您可能更关注正确预测的正响应的百分比。

- **最小可能性。** 将得分值大于指定值的记录指定到分类表中的预测正响应类别。由过程生成的得分代表联系人将做出正面响应（例如，购买产品）的可能性。作为一般规则，您必须指定一个接近最低目标响应率的比例值。例如，如果您对至少 5% 的响应率感兴趣，请指定 0.05。值必须大于 0 且小于 1。

重新编码的响应字段的名称和标签

该过程自动将响应字段重新编码为新字段，其中 1 代表正响应，0 代表负响应，并在重新编码的字段上执行分析。您可以使用自己的名称和标签覆盖默认名称和标签。名称必须符合 IBM® SPSS® Statistics 命名规则。

保存得分

在原始数据集中自动保存一个包含倾向得分的新字段。得分代表正响应的可能性，并以比例表示。

- 字段名必须符合 SPSS Statistics 命名规则。
- 该字段名不能与数据集中现有字段名重复。如果在同一数据集上多次运行此过程，则需要每次指定不同的名称。

创建分类响应字段

响应字段应当为分类字段，且具有一个代表所有正响应的值。任何其他非缺失值均被假设为负响应。如果响应字段代表连续（尺度）值，例如购买数量或购买金额，则需要创建新字段，以便为所有非零响应值指定单个正响应值。

- 如果负响应被记录为 0（不是空值，后者作为缺失处理），则可以通过以下公式进行计算：

  NewName=OldName>0

  其中 NewName 为新字段的名称，OldName 为原始字段的名称。这是一个逻辑表达式，它为所有大于零的非缺失值指定值 1，为所有小于或等于零的非缺失值指定值 0。
- 如果未记录有负响应值，这些值将作为缺失处理，公式则更为复杂：

  NewName=NOT(MISSING(OldName))

在此逻辑表达式中，为所有非缺失响应值指定值 1，为所有缺失响应值指定值 0。

- 如果不能区分负 (0) 响应值和缺失值，则无法计算准确的响应值。如果实际缺失值相对较少，这对计算的响应率可能并无显著影响。但如果缺失值较多，比如当仅为整个数据集中少量检验样本记录响应信息时，则计算的响应率将没有意义，因为它们将明显低于实际响应率。

创建分类响应字段

▶ 从菜单中选择：
转换 > 计算变量

▶ 为“目标变量”输入新的字段（变量）名称。

▶ 如果负响应被记录为 0，则为“数值表达式”输入 **OldName>0**，其中 OldName 为原始字段名。

▶ 如果负响应被记录为缺失（空），则为“数值表达式”输入 **NOT(MISSING(OldName))**，其中 OldName 为原始字段名。

章
7

控制包装检验

该方法比较市场营销活动，以检查不同包装或商品之间是否存在显著的效果差异。活动效果通过响应来测量。“活动”字段标识不同的活动，例如，Offer A 和 Offer B。“响应”字段指示联系人对活动有无响应。在响应被记录为购买金额（例如“99.99”）时选择“购买金额”。在响应只是指示联系人是否正面回应（例如“是”或“否”）时，选择“回应”。

示例。 公司直销部门想了解新的包装设计能否产生比现有包装更多的正面响应。因此他们发出测试邮件，以确定新包装能否产生明显更高的正响应率。测试邮件包括获得现有包装的控制组和获得新包装设计的测试组。然后比较两组的结果，看看是否存在显著差异。

输出

输出包含两个表格，其中一个显示由活动字段定义的每个组的正、负响应计数与百分比，另一个则标识存在明显差异的组。

图片 7-1
控制包装检验输出

		控制包装			
		控制		检验	
		计数	列 N %	计数	列 N %
效果 (1=是 0=否)	0	879	96.6%	984	97.7%
	1	31	3.4%	23	2.3%

在 控制 和 检验 之间不存在统计显著性差异。

控制包装检验数据注意事项和假设

活动字段。 “活动”字段必须为分类字段（名义或有序）。

效果响应字段。 如果选择“购买金额”作为效果字段，则此字段必须为数值，且测量级别应当为连续（尺度）。

如果不能区分负（对于购买金额，则为 0 值）响应值和缺失值，则无法计算准确的响应率。如果实际缺失值相对较少，这对计算的响应率可能并无显著影响。但如果缺失值较多，比如当仅为整个数据集中少量检验样本记录响应信息时，则计算的响应率将没有意义，因为它们将明显低于实际响应率。

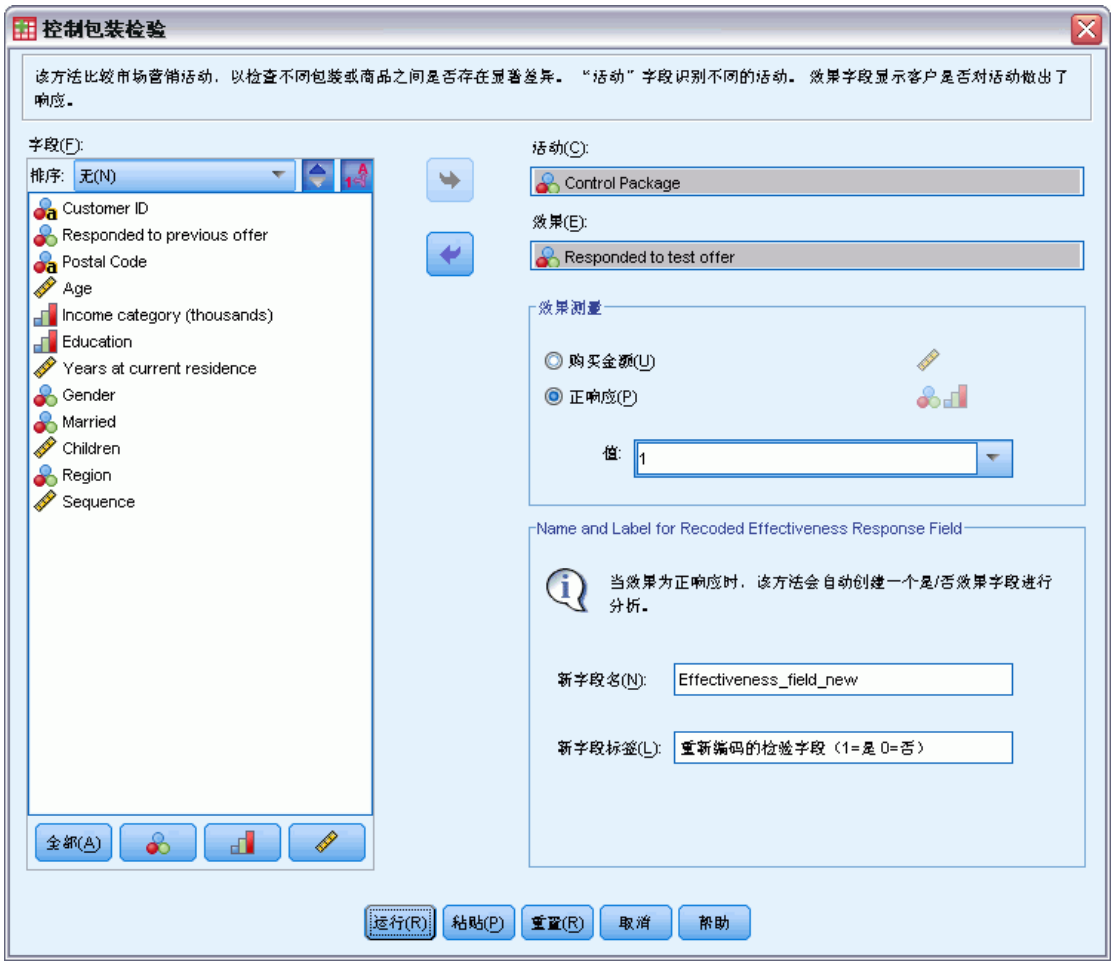
假设。 此过程假设联系人被随机分配到每个活动组。换句话说，不存在特定人口统计学、购买历史或其他特征会影响组分配，所有联系人以相同概率分配到任意组。

获取控制包装检验

从菜单中选择：
直销 > 选择方法

▶ 选择比较活动效果。

图片 7-2
“控制包装检验”对话框



▶ 选择标识每个联系人所属活动组的字段（例如 Offer A、Offer B 等）。此字段必须为名义或有序字段。

▶ 选择指示响应效果的字段。

如果响应字段为购买金额，则此字段必须为数值。

如果响应字段仅指示联系人是否正面回应（例如，“是”或“否”），则选择**回应**并输入代表正响应的值。如果有任何值定义了值标签，则可以从下拉列表中选择值标签，这将显示对应的值。

将自动创建新字段，其中 1 代表正响应，0 代表负响应，并在此新字段上执行分析。您可以使用自己的名称和标签覆盖缺省名称和标签。名称必须符合 IBM® SPSS® Statistics 命名规则。

▶ 单击**运行**以运行该过程。

部分 II:
示例

章 8

交易数据的 RFM 分析

在交易数据文件中，每行代表一笔单独的交易，而非单独的客户，每个客户可能有多个交易行。本示例使用数据文件 rfm_transactions.sav。 有关详细信息，请参阅第 90 页码附录 A 中的样本文件。

交易数据

数据集必须包含含有以下信息的变量：

- 标识每个个案（客户）的变量或变量组合。
- 拥有每次交易日期的变量。
- 拥有每次交易的消费金额值的变量。

图片 8-1
RFM 交易数据

ID	Date	Amount
1	08/04/2005	129
1	10/25/2004	50
1	07/24/2004	118
1	07/24/2004	136
1	09/04/2006	52
2	09/23/2005	183
2	11/05/2004	24
2	11/10/2005	66
2	12/03/2004	77
3	06/04/2005	102
3	05/15/2005	131

运行分析

▶ 要计算 RFM 得分，从菜单中选择：
直销 > 选择方法

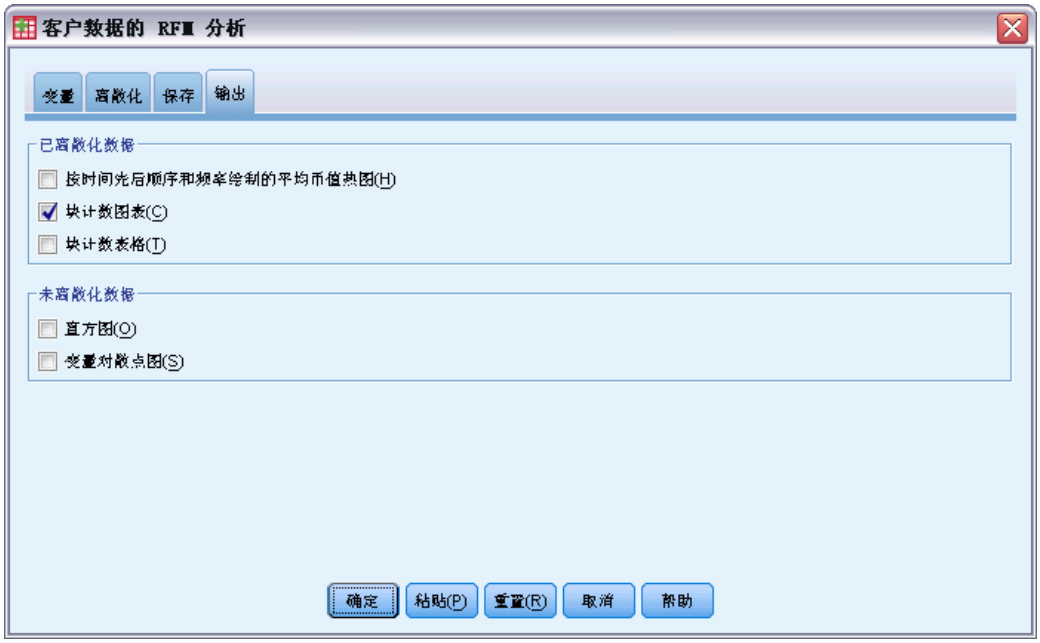
▶ 选择**帮助标识我的最佳联系人（RFM 分析）**，并单击**继续**。

▶ 单击**交易数据**，然后单击**继续**。

图片 8-2
交易数据的 RFM，“变量”选项卡

- 单击**重置**清除以前的所有设置。
- 对于“交易数据”，选择购买日期 [Date]。
- 对于“交易金额”，选择购买金额 [Amount]。
- 对于“摘要方法”，选择**总计**。
- 对于“客户标识”，选择客户 ID [ID]。
- 然后单击**输出**选项卡。

图片 8-3
交易数据的 RFM，“输出”选项卡



- 选择（选中）**块计数图表**。
- 然后单击**确定**以运行该过程。

评估结果

在从交易数据计算 RFM 得分时，将创建包含新的 RFM 得分的新数据集。

图片 8-4
交易数据集的 RFM

ID	Date_most_Recent	Transaction_count	Amount	Recency_score	Frequency_score	Monetary_score	RFM_score
1	05/17/2006	10	1313.00	2	3	5	235
2	09/21/2005	11	1230.00	1	5	4	154
3	08/11/2006	13	1194.00	3	5	2	352
4	05/24/2006	9	794.00	2	3	2	232
5	03/13/2005	3	278.00	1	1	2	112
6	07/28/2006	9	922.00	3	2	4	324
7	06/20/2006	11	961.00	2	4	2	242

默认情况下，该数据集包含每个客户的下列信息：

- “客户 ID”变量
- 最近交易日期
- 交易总数

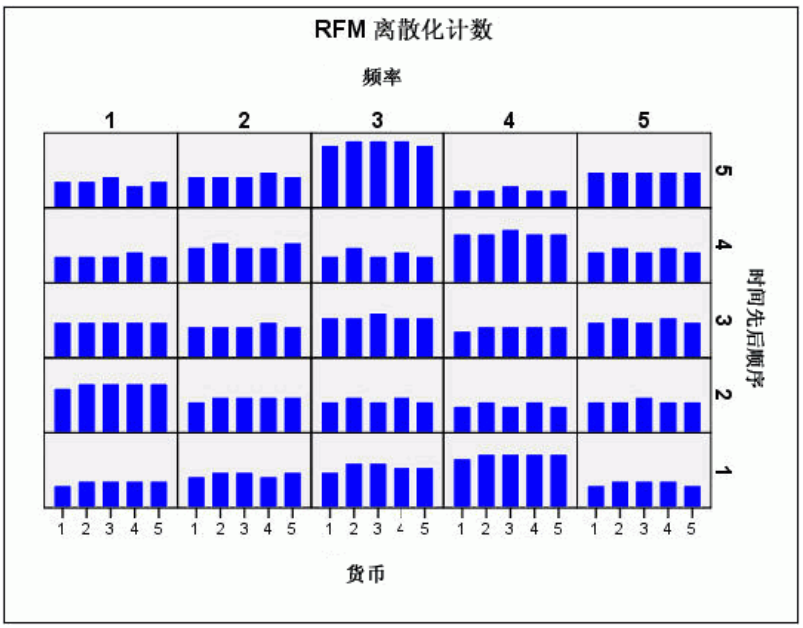
- 交易摘要金额（默认为总计）。
- 崭新、频率、消费金额和合并 RFM 得分

新数据集中仅为每个客户包含一行数据（一条记录）。原始交易数据已按客户标识变量值进行汇总。标识变量始终包含在新数据集中，否则无法将 RFM 得分匹配到客户。

每个客户的合并 RFM 得分由三个独立得分简单拼接而成，计算方法为：(崭新得分 x 100) + (频率得分 x 10) + 消费金额得分。

查看器窗口中的块计数图表显示每个 RFM 类别中的客户数）。

图片 8-5
块计数图表



默认方法为对 3 个 RFM 组成要素分别应用 5 个得分类别，这会产生 125 个可能的 RFM 得分类别。图表中的每个条代表每个 RFM 类别中的客户数。

理想情况下，客户应当在所有 RFM 得分类别之间相对均匀地分布。但实际通常存在一定的偏差，如本例中所示。如果存在许多空类别，则可能需要考虑更改离散化方法。

有多种处理 RFM 得分非均匀分布的策略，包括：

- 使用嵌套代替独立离散化。
- 减少可能的得分类别（块）数。
- 如果存在大量同数的值，则将具有相同得分的个案随机分配到不同类别。

有关详细信息，请参阅第 6 页码第 2 章中的RFM 离散化。

合并“得分数据”与“客户数据”

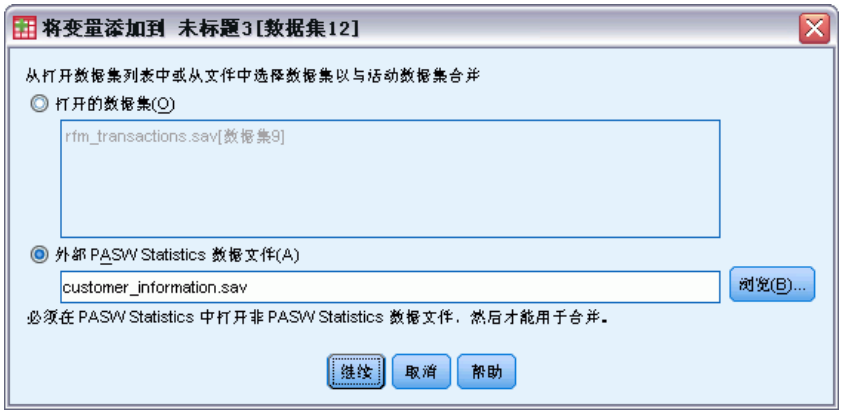
现在，已经有了包含 RFM 得分的数据集，还需要将这些得分匹配到客户。可以将得分合并回交易数据文件，但通常需要将得分数据与某个为每个客户包含一行数据（一条记录）且同时包含客户姓名、地址等信息的数据文件（如 RFM 得分数据集）进行合并。

图片 8-6
“变量视图”中的 RFM 得分数据集

名称	类型	宽度	小数	标签	值
ID	数值(N)	5	0	Customer ID	无
Date_most_Recent	日期	10	0	Date of most rec...	无
Transaction_count	数值(N)	7	0	Number of trasa...	无
Amount	数值(N)	8	2	Amount	无
Recency_score	数值(N)	3	0	Recency score	无
Frequency_score	数值(N)	3	0	Frequency score	无
Monetary_score	数值(N)	3	0	Monetary score	无
RFM_score	数值(N)	3	0	RFM score	无

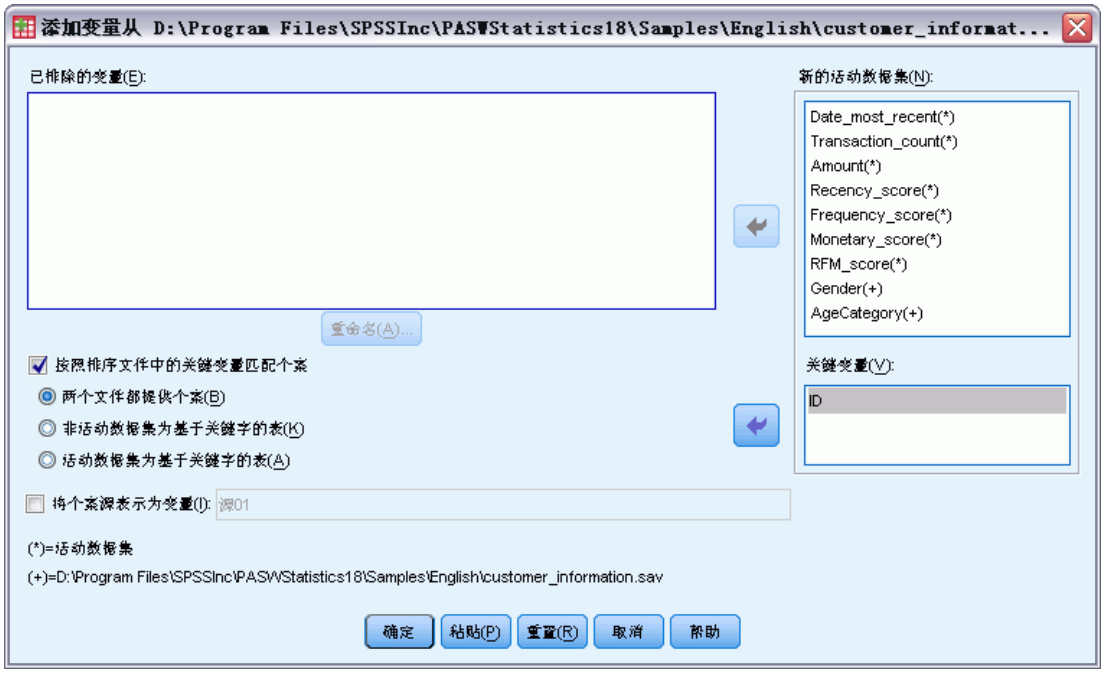
- 使包含 RFM 得分的数据集成为活动数据集。（在包含该数据集的“数据编辑器”窗口中单击任意位置。）
- 从菜单中选择：
  数据 > 合并文件 > 添加变量

图片 8-7
添加变量，“选择文件”对话框



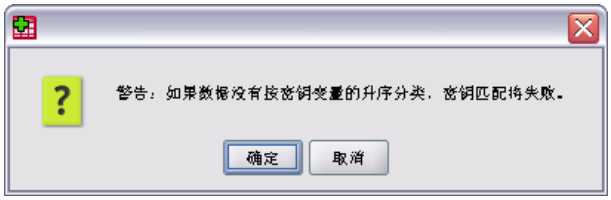
- 选择**外部数据文件**。
- 使用**浏览**按钮导航到 示例 文件夹，并选择 customer_information.sav。有关详细信息，请参阅第 90 页码附录 A 中的样本文件。
- 然后单击**继续**。

图片 8-8
添加变量，“选择变量”对话框



- 选择（选中）**按照排序文件中的关键变量匹配个案**。
- 选择**两个文件都提供个案**。
- 选择“关键变量”列表的 ID。
- 单击**确定**。

图片 8-9
“添加变量”警告消息



请注意用于提醒您两个文件都必须按关键变量的升序进行排列的警告消息。在本例中，两个文件已按关键变量的升序进行排列，此处的关键变量是计算 RFM 得分时所选的客户标识变量。在从交易数据计算 RFM 得分时，新数据集自动按客户标识变量升序进行排列。如果更改了得分数据集的排列顺序，或打算用来与得分数据集合并的数据文件是以其他顺序排列，则必须首先按客户标识变量的升序来排列这两个文件。

- 单击**确定**以合并两个数据集。

包含 RFM 得分的数据集现在也包含了每个客户的姓名、地址及其他信息。

图片 8-10
合并的数据集

名称	类型	宽度	小数	标签	值
ID	数值(N)	5	0	Customer ID	无
Date_most_Recent	日期	10	0	Date of most rec...	无
Transaction_count	数值(N)	7	0	Number of trasa...	无
Amount	数值(N)	8	2	Amount	无
Recency_score	数值(N)	3	0	Recency score	无
Frequency_score	数值(N)	3	0	Frequency score	无
Monetary_score	数值(N)	3	0	Monetary score	无
RFM_score	数值(N)	3	0	RFM score	无
Name	字符串	4	0		无
Address	字符串	7	0		无
City	字符串	4	0		无
State_Province	字符串	14	0		无
Postalcode	字符串	11	0		无
Country	字符串	7	0		无
Gender	数值(N)	1	0		{0, 女}...
AgeCategory	数值(N)	1	0	Age Category	{1, <25}...

章 9

聚类分析

聚类分析是用于揭示数据中的自然分组（或聚类）的探索性工具。例如，它可以根据各种人口统计和购买特征识别不同的客户组。

例如，公司直销部门想要标识其客户数据库中的人口统计组，以帮助确定市场营销活动策略和开发新产品。

这些信息收集在 dmdata.sav 中。 有关详细信息，请参阅第 90 页码附录 A 中的样本文件。

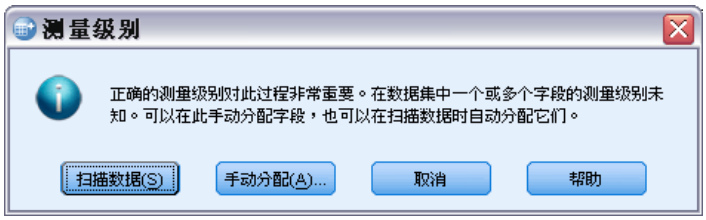
运行分析

▶ 要运行“聚类分析”，请从菜单中选择：
直销 > 选择方法

▶ 选择**将我的联系人分段到聚类**并单击**继续**。

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

图片 9-1
测量级别警报



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

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

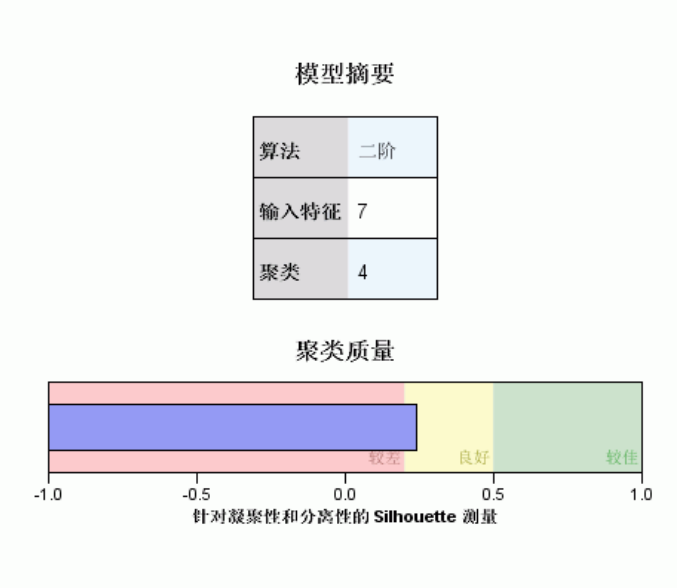
在本示例文件中，不存在具有未知测量级别的字段，并且所有字段均具有正确的测量级别；因此不会显示测量级别警报。

图片 9-2
“聚类分析：字段”选项卡

- 选择以下字段以创建段：年龄、收入类别、教育、当前地址居住年限、性别、已婚和孩子。
- 单击运行以运行该过程。

输出

图片 9-3
聚类模型摘要

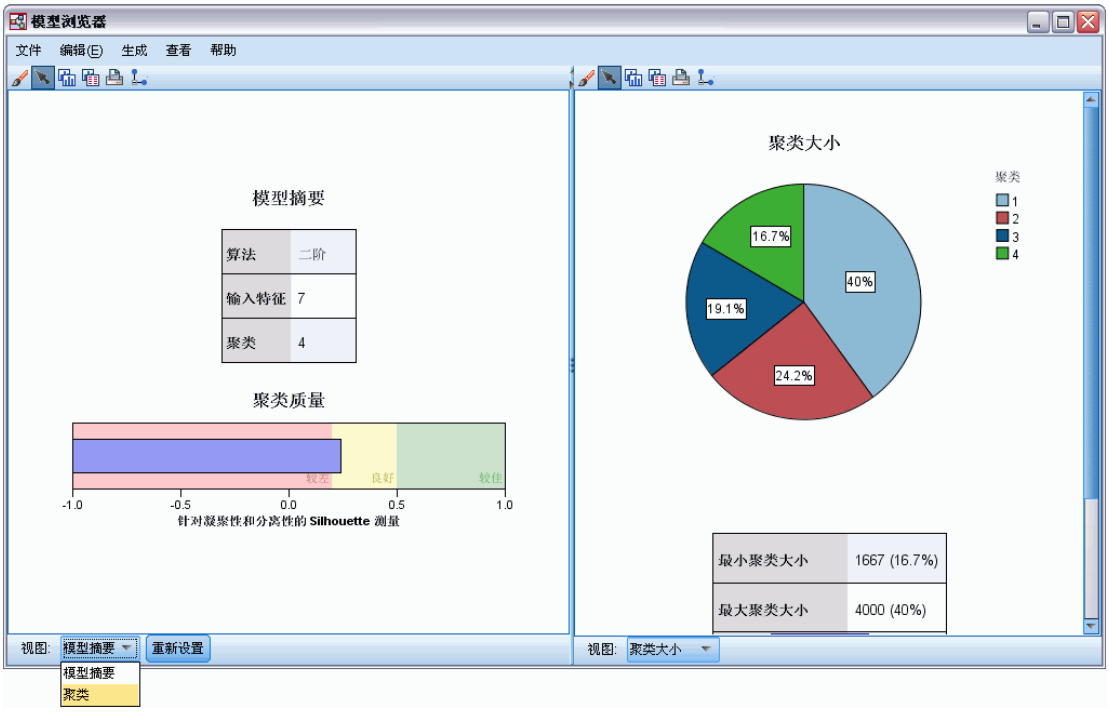


结果显示在“聚类模型查看器”中。

- 模型摘要显示根据您选择的 7 个输入特征（字段）找到了四个聚类。
- 聚类质量图表显示整体模型质量在“良好”范围中间。

▶ 双击“聚类模型查看器”输出以激活“模型查看器”。

图片 9-4
激活的聚类模型浏览器



▶ 从“聚类模型查看器”窗口底部的“视图”下拉列表中选择**聚类**。

图片 9-5
聚类视图

聚类	1	2	3	4
标签				
描述				
大小	40.0% (4000)	24.2% (2424)	19.1% (1909)	16.7% (1667)
特征	Age 50.30	Age 44.07	Age 39.05	Age 33.09
	Children 1.58	Children 1.29	Children 0.39	Children 0.12
	Gender Male (57.0%)	Gender Female (100.0%)	Gender Male (100.0%)	Gender Female (50.9%)
	Income category (thousands) 75+ (56.1%)	Income category (thousands) 50-74 (47.2%)	Income category (thousands) 75+ (34.8%)	Income category (thousands) <25 (100.0%)
	Married Yes (100.0%)	Married No (78.5%)	Married No (100.0%)	Married No (78.5%)
	Education Post-graduate (20.5%)	Education Post-graduate (20.5%)	Education College (21.1%)	Education Post-graduate (20.6%)
	Years at current residence 9.47	Years at current residence 9.51	Years at current residence 9.47	Years at current residence 9.42

“聚类”视图显示每个聚类的属性信息。

- 对于连续（尺度）字段，显示均值（平均值）。
- 对于分类（名义、有序）字段，显示众数。众数是具有最多记录数的类别。在本例中，每个记录是一个客户。
- 默认情况下，字段按其对模型的整体重要性顺序显示。在本例中，年龄具有最高的整体重要性。您也可以按聚类内重要性或字母顺序排列字段。

如果您在“聚类”视图中选中（单击）任何单元格，您可以看到汇总该聚类字段值的图表。

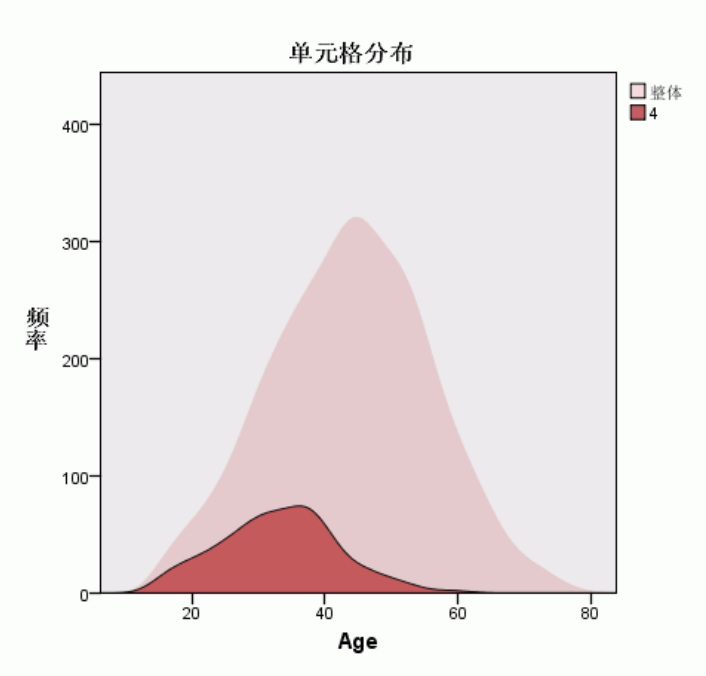
▶ 例如，选中聚类 1 的年龄单元格。

图片 9-6
聚类 1 的年龄直方图

对于连续字段，显示直方图。直方图显示该聚类中值的分布以及字段值的整体分布。直方图表明聚类 1 中的客户年龄较大。

▶ 在“聚类”视图中选中聚类 4 的年龄单元格。

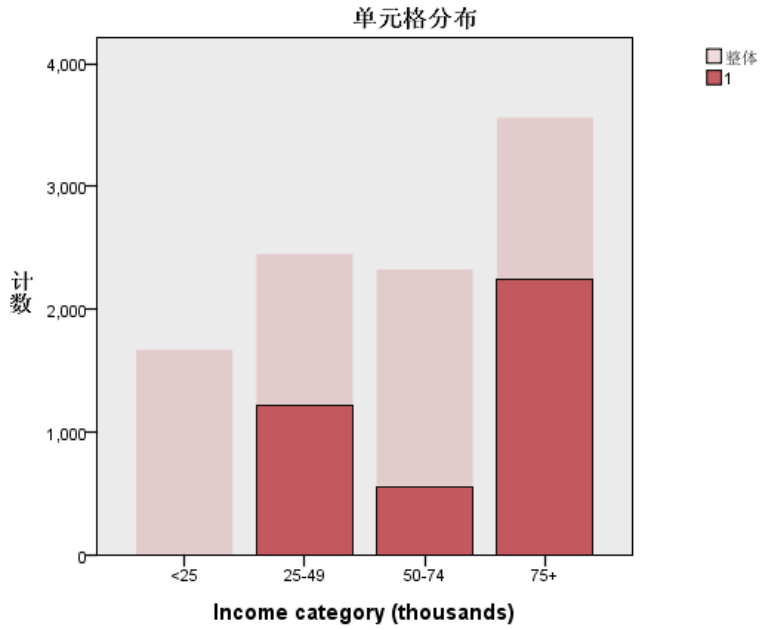
图片 9-7
聚类 4 的年龄直方图



与聚类 1 相反，聚类 4 中的客户年龄小于整体平均年龄。

▶ 在“聚类”视图中选中聚类 1 的收入类别单元格。

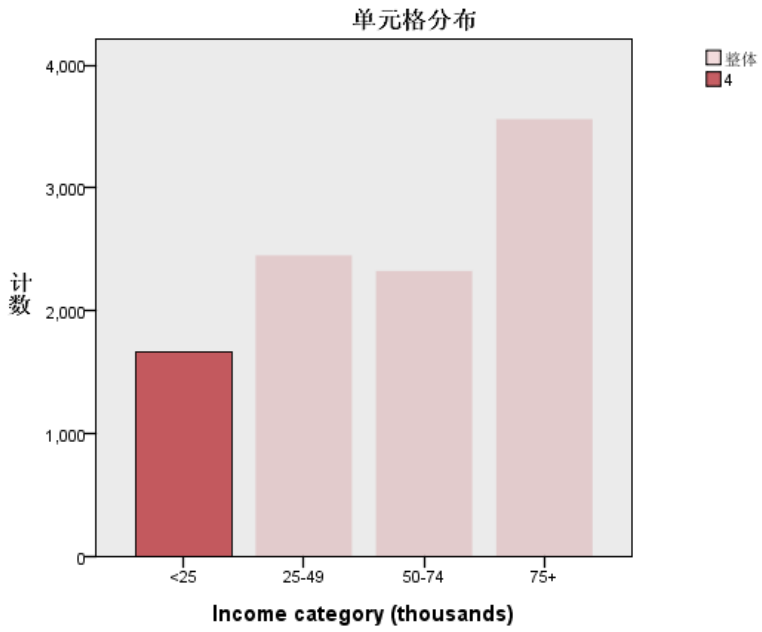
图片 9-8
聚类 1 的收入类别条形图



对于分类字段，显示条形图。在该聚类的收入类别条形图中，最值得注意的特征是在最低收入类别中不存在任何客户。

▶ 在“聚类”视图中选中聚类 4 的收入类别单元格。

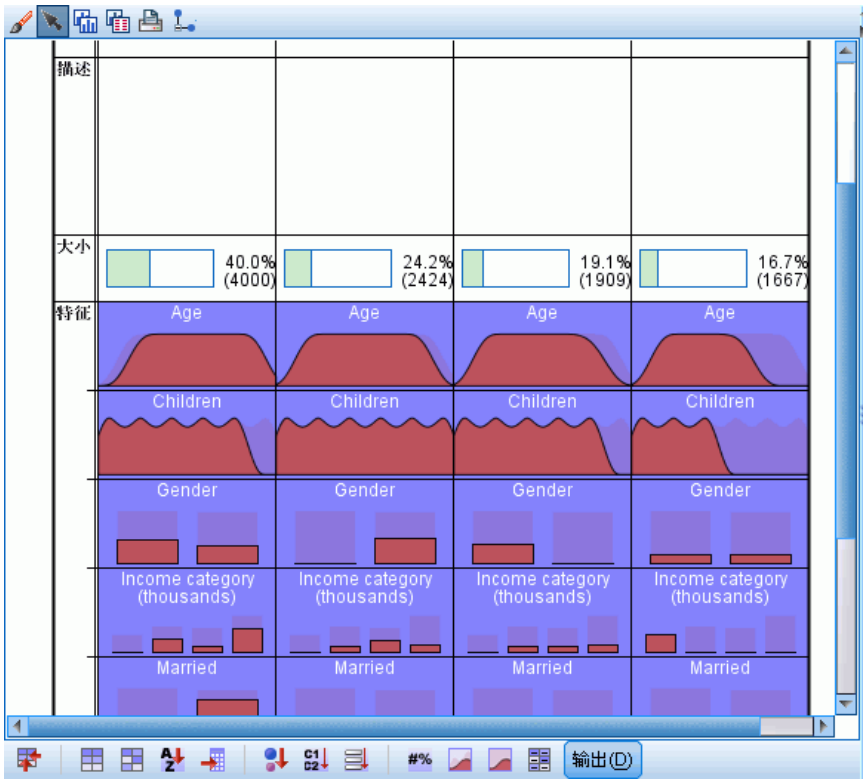
图片 9-9
聚类 4 的收入类别条形图



与聚类 1 相反，聚类 4 中的所有客户都在最低收入类别中。

您也可以通过“模型查看”窗口底部的工具栏更改“聚类”视图，以便在单元格中显示图表，这样可以快速比较聚类间的值分布。

图片 9-10
聚类中显示的图表



查看“聚类”视图和图表中提供的每个单元格的附加信息，您可以看到聚类间的某些不同差异：

- 聚类 1 中的客户多数年龄较大，已婚且有小孩，收入较高。
- 聚类 2 中的客户多为年龄较大的单身母亲，且收入适中。
- 聚类 3 中的客户多为年轻的单身男士，没有小孩。
- 聚类 4 中的客户多为年轻的单身女士，没有小孩且收入较低。

“聚类”视图中的“描述”单元格是文本字段，您可编辑以添加每个聚类的描述。

图片 9-11
具有聚类描述的聚类视图

聚类	1	2	3	4
标签				
描述	Older, married, have children, higher income	Older single mothers, moderate income	Younger single men, no children	Younger single women, no children, low income
大小	40.0% (4000)	24.2% (2424)	19.1% (1909)	16.7% (1667)
特征	Age 50.30	Age 44.07	Age 39.05	Age 33.09
	Children 1.58	Children 1.29	Children 0.39	Children 0.12
	Gender Male (57.0%)	Gender Female (100.0%)	Gender Male (100.0%)	Gender Female (50.9%)
	Income category (thousands) 75+ (56.1%)	Income category (thousands) 50-74 (47.2%)	Income category (thousands) 75+ (34.8%)	Income category (thousands) <25 (100.0%)
	Married Yes (100.0%)	Married No (78.5%)	Married No (100.0%)	Married No (78.5%)
	Education Post-graduate (20.5%)	Education Post-graduate (20.5%)	Education College (21.1%)	Education Post-graduate (20.6%)
	Years at current residence 9.47	Years at current residence 9.51	Years at current residence 9.47	Years at current residence 9.42

根据聚类选择记录

您可使用以下两种方法来根据聚类成员选择记录：

- 在“聚类模型查看器”中以交互方式创建过滤条件。
- 使用由过程生成的聚类字段值指定过滤或选择条件。

在聚类模型查看器中创建过滤条件

要在“聚类模型查看器”中创建从特定聚类中选择记录的过滤条件：

1. 激活（双击）“聚类模型浏览器”。
2. 从“聚类模型查看器”窗口底部的“视图”下拉列表中选择**聚类**。

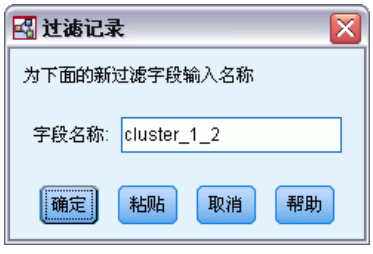
▶ 在“聚类视图”顶部单击你需要的聚类的聚类编号。如果您想选择多个聚类，Ctrl-单击每个您需要的附加聚类编号。

图片 9-12
在“聚类”视图中选中的聚类

聚类	1	2	3	4
标签				
描述	Older, married, have children, higher income	Older single mothers, moderate income	Younger single men, no children	Younger single women, no children, low income
大小	40.0% (4000)	24.2% (2424)	19.1% (1909)	16.7% (1667)
特征	Age 50.30	Age 44.07	Age 39.05	Age 33.09
	Children 1.58	Children 1.29	Children 0.39	Children 0.12
	Gender Male (57.0%)	Gender Female (100.0%)	Gender Male (100.0%)	Gender Female (50.9%)
	Income category (thousands) 75+ (56.1%)	Income category (thousands) 50-74 (47.2%)	Income category (thousands) 75+ (34.8%)	Income category (thousands) <25 (100.0%)
	Married Yes (100.0%)	Married No (78.5%)	Married No (100.0%)	Married No (78.5%)
	Education Post-graduate (20.5%)	Education Post-graduate (20.5%)	Education College (21.1%)	Education Post-graduate (20.6%)
	Years at current residence 9.47	Years at current residence 9.51	Years at current residence 9.47	Years at current residence 9.42

▶ 从“聚类模型查看器”菜单中选择：
生成 > 过滤记录

图片 9-13
“过滤记录”对话框



▶ 为过滤字段输入名称并单击**确定**。名称必须符合 IBM® SPSS® Statistics 命名规则。

图片 9-14
“数据编辑器”中的已过滤记录

	ID	Married	Children	Region	ClusterGroup1	clusters_1_2
~~14~~	03623	No	0	West	3	.00
~~15~~	01353	No	0	West	3	.00
~~16~~	07055	No	0	West	3	.00
17	04455	No	0	West	2	1.00
18	07210	No	1	West	2	1.00
~~19~~	08054	No	0	West	4	.00
~~20~~	06937	No	0	West	4	.00
~~21~~	06512	No	0	West	4	.00
~~22~~	08315	No	0	West	4	.00
23	09676	No	3	West	2	1.00
~~24~~	09636	No	0	West	4	.00
25	08579	No	1	West	2	1.00
26	01480	No	1	West	2	1.00

这会在数据集中创建一个新字段并根据该字段的值过滤数据集中的记录。

- 过滤字段值为 1 的记录将被包括在后续分析、图表和报表中。
- 过滤字段值为 0 的记录将被排除。
- 排除的记录不会从数据集中删除。它们保留了过滤状态指示符，该指示符在“数据编辑器”中以贯穿记录号的对角线表示。

根据聚类字段值选择记录

默认情况下，“聚类分析”会创建一个新字段用来标识每个记录的聚类组。该字段的默认名称为 ClusterGroupn，其中 n 是一个整数，它构成了独有的字段名。

图片 9-15
添加到数据集的聚类字段

	ID	Gender	Married	Children	Region	ClusterGroup1
1	01359	Female	No	0	West	4
2	06262	Female	No	1	West	2
3	08031	Male	No	0	West	3
4	01971	Male	No	0	West	4
5	09689	Male	No	0	West	3
6	06108	Male	No	1	West	3
7	09853	Male	No	0	West	3
8	06802	Male	No	0	West	4
9	07597	Male	No	0	West	3
10	03692	Male	No	1	West	3
11	00071	Male	No	0	West	4
12	00769	Male	No	0	West	3

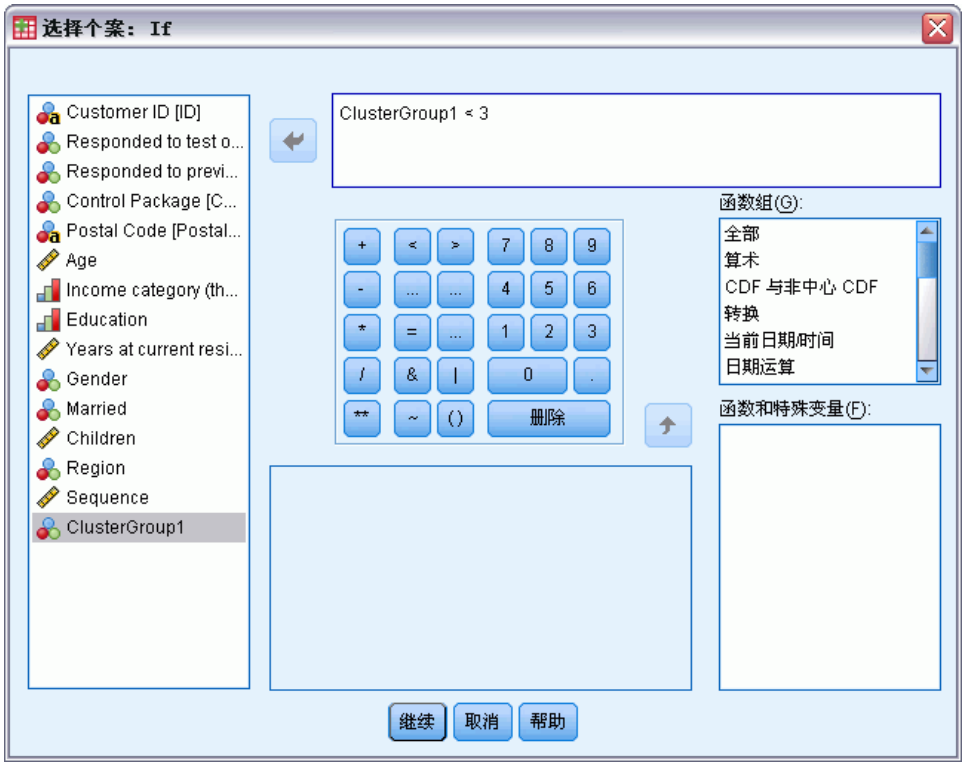
要使用聚类字段的值选择特定聚类中的记录：

▶ 从菜单中选择：
数据 > 选择个案

图片 9-16
“选择个案”对话框

▶ 在“选择个案”对话框中，选择**如果条件满足**，然后单击**如果**。

图片 9-17
选择个案：“如果”对话框



▶ 输入选择条件。

例如，ClusterGroup1 < 3 将选择聚类 1 和 2 中的所有记录，同时排除聚类 3 和更高编号的聚类中的记录。

▶ 单击继续。

在“选择个案”对话框中，有几个与如何处理已选定和未选定记录相关的选项：

过滤掉未选定的个案。 这会创建一个指定过滤条件的新字段。排除的记录不会从数据集中删除。它们保留了过滤状态指示符，该指示符在“数据编辑器”中以贯穿记录号的对角线表示。这等同于在“聚类模型查看器”中以交互方式选择聚类。

将选定的个案复制到新数据集。 这会在当前会话中创建一个只包含满足过滤条件的记录的新数据集。原始数据集不受影响。

删除未选定个案。从数据集删除未选定记录。只有退出文件而不保存任何更改，然后重新打开文件，才能恢复删除的记录。如果保存对数据文件的更改，则会永久删除个案。

“选择个案”对话框还提供了将现有变量用作过滤变量（字段）的选项。如果您在“聚类模型查看器”中以交互方式创建过滤条件并将生成的过滤字段随数据集一起保存，则您可以在后续会话中使用该字段过滤记录。

摘要

“聚类分析”是有用的探索性工具，它可以揭示数据中的自然分组（或聚类）。您可以使用这些聚类的信息确定市场营销活动策略和开发新产品。您可以根据聚类成员选择记录用于进一步分析或目标市场营销活动。

章
10

潜在客户概要文件

潜在客户概要文件使用先前或检验活动的结果来创建描述概要文件。您可以使用概要文件在未来的活动中集中面向特定的联系人群体。例如，根据测试邮件的结果，公司直销部门想要生成以人口统计信息为基础的最可能对某种类型的产品做出响应的人员类型概要文件。根据这些结果，就可确定针对类似产品应当使用哪些邮寄列表类型。

例如，公司直销部门向其总客户数据库中约 20% 的客户发送测试邮件。测试邮件结果记录在数据文件中，该数据文件还包含每个客户的人口统计特征，包括年龄、性别、婚姻状况和地理区域。结果则以简单的是/否形式进行记录，表示测试邮件中哪些客户已响应（购买），哪些未响应。

这些信息收集在 dmdata.sav 中。 有关详细信息，请参阅第 90 页码附录 A 中的样本文件。

数据注意事项

响应字段应当为分类字段，且具有一个代表所有正响应的值。任何其他非缺失值均被假设为负响应。如果响应字段代表连续（尺度）值，例如购买数量或购买金额，则需要创建新字段，以便为所有非零响应值指定单个正响应值。有关详细信息，请参阅第 21 页码第 4 章中的创建分类响应字段。

运行分析

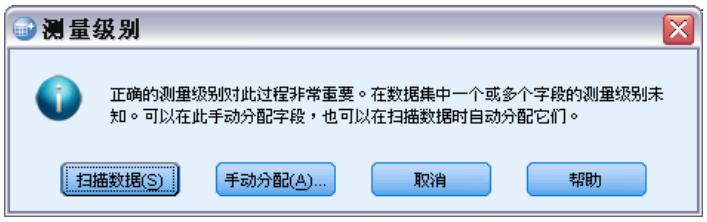
- 要运行潜在客户概要文件分析，从菜单中选择：

  直销 > 选择方法

- 选择**生成对产品做出响应的我的联系人的概要文件**，然后单击**继续**。

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

图片 10-1
测量级别警报



- **扫描数据。** 读取活动数据集中的数据，并分配默认测量级别给任何具有当前未知测量级别的字段。如果数据集较大，该过程可能需要一些时间。

- **手动分配。** 打开列出了所有具有未知测量级别的字段的对话框。您可以使用该对话框将测量级别分配给这些字段。您也可以在数据编辑器的变量视图中分配测量级别。

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

在本示例文件中，不存在具有未知测量级别的字段，并且所有字段均具有正确的测量级别；因此不会显示测量级别警报。

图片 10-2
“潜在客户概要文件：字段”选项卡

- 对于响应字段，选择已对测试产品做出响应。
- 对于正响应值，从下拉列表中选择是。在文本字段中显示值 1，因为“是”实际上为与记录值 1 关联的值标签。（如果正响应值未定义有值标签，您只需在文本字段中输入值即可。）
- 对于“创建概要文件”，选择年龄、收入类别、教育、当前地址居住年限、性别、已婚、地区和孩子。
- 单击**设置**选项卡。

图片 10-3
“潜在客户概要文件：设置”选项卡

- 选择（选中）“在结果中包括最小响应率阈值信息”。
- 对于目标响应率，输入 7。
- 然后单击运行以运行该过程。

输出

图片 10-4
响应率表格

数字	概要文件			
	描述	组大小	响应率	累积响应率
1	Region = "West","South","East" Gender = "Female" Married = "No"	379	9.2%	9.2%
2	Region = "West","South","East" Gender = "Female" Married = "Yes"	299	5.0%	7.4%
3	Region = "West","South","East" Gender = "Male"	722	4.7%	6.0%
4	Region = "North"	517	2.5%	5.1%

绿色: 满足目标响应率。
红色: 不满足目标响应率。

响应率表格显示过程标识的每个概要文件组的信息。

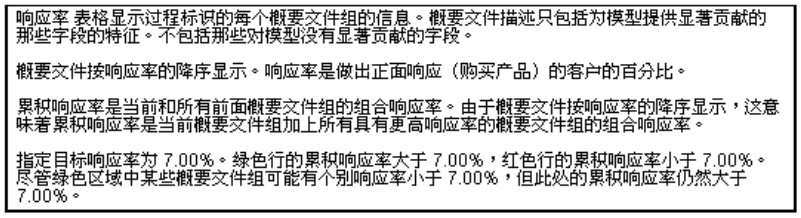
- 概要文件按响应率的降序显示。
- 响应率是做出正面响应（购买产品）的客户的百分比。
- 累积响应率是当前和所有先前概要文件组的组合响应率。由于概要文件按响应率的降序显示，这意味着累积响应率是当前概要文件组加上所有具有更高响应率的概要文件组的组合响应率。
- 概要文件描述只包括那些为模型提供显著贡献的字段的特征。在本例中，模型中包括地区、性别和婚姻状况。而其余字段，年龄、收入、教育和当前地址居住年限，则未包括在内，因为它们对模型没有显著贡献。
- 表格的绿色区域代表其累积响应率等于或大于指定目标响应率（在本例中为 7%）的概要文件组。
- 表格的红色区域代表其累积响应率低于指定目标响应率的概要文件组。
- 表格最后一行中的累积响应率是测试邮件中包括的所有客户的整体或平均响应率，因为它是所有概要文件组的响应率。

表中所示的结果表明，如果目标女性在西部、南部和东部，则应当获得比目标响应率稍高的响应率。

然而，请注意，在这些地区中未婚女性 (9.2%) 和已婚女性 (5.0%) 的响应率之间存在显著差异。尽管两个组的累积响应率均高于目标响应率，但后一组的响应率实际上却低于目标响应率，这表明您可以寻找其他可以改进模型的特征。

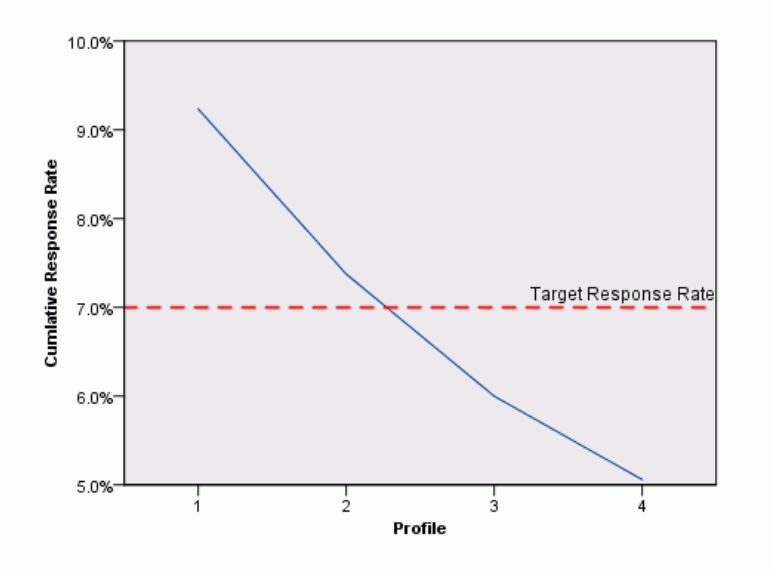
智能输出

图片 10-5
智能输出



此表带有“智能输出”功能，用以提供有关如何解释表的一般信息，以及表中所包含结果的特定信息。

图片 10-6
累积响应率图表



累积响应率图表基本上是表格中显示的累积响应率的可视化表示。由于概要文件是按响应率的降序报告，因此对于每个后续概要文件而言累积响应率行始终在下降。与表格类似，图表显示在概要文件组 2 和概要文件组 3 之间累积响应率下降到低于目标响应率的位置。

摘要

在此次特定的测试邮件活动中，标识了四个概要文件组，结果显示似乎只有性别、地区和婚姻状况这三个显著人口统计特征与某个人是否对产品做出响应有关。其中居住在南部、东部和西部的未婚女性组成了最高响应率组。此后响应率快速下降，尽管在相同地区中的已婚女性仍然取得比目标响应率更高的累积响应率。

章
11

邮政编码响应率

此方法使用先前活动的结果来计算邮政编码响应率。这些响应率可以用于在未来的活动中集中面向特定的邮政编码。

例如，根据先前邮件的结果，公司直销部门按邮政编码生成响应率。然后，根据不同的标准，例如最低可接受响应率和/或在邮件中包括的最大联系人数量，他们可以集中面向特定的邮政编码。

这些信息收集在 dmdata.sav 中。 有关详细信息，请参阅第 90 页码附录 A 中的样本文件。

数据注意事项

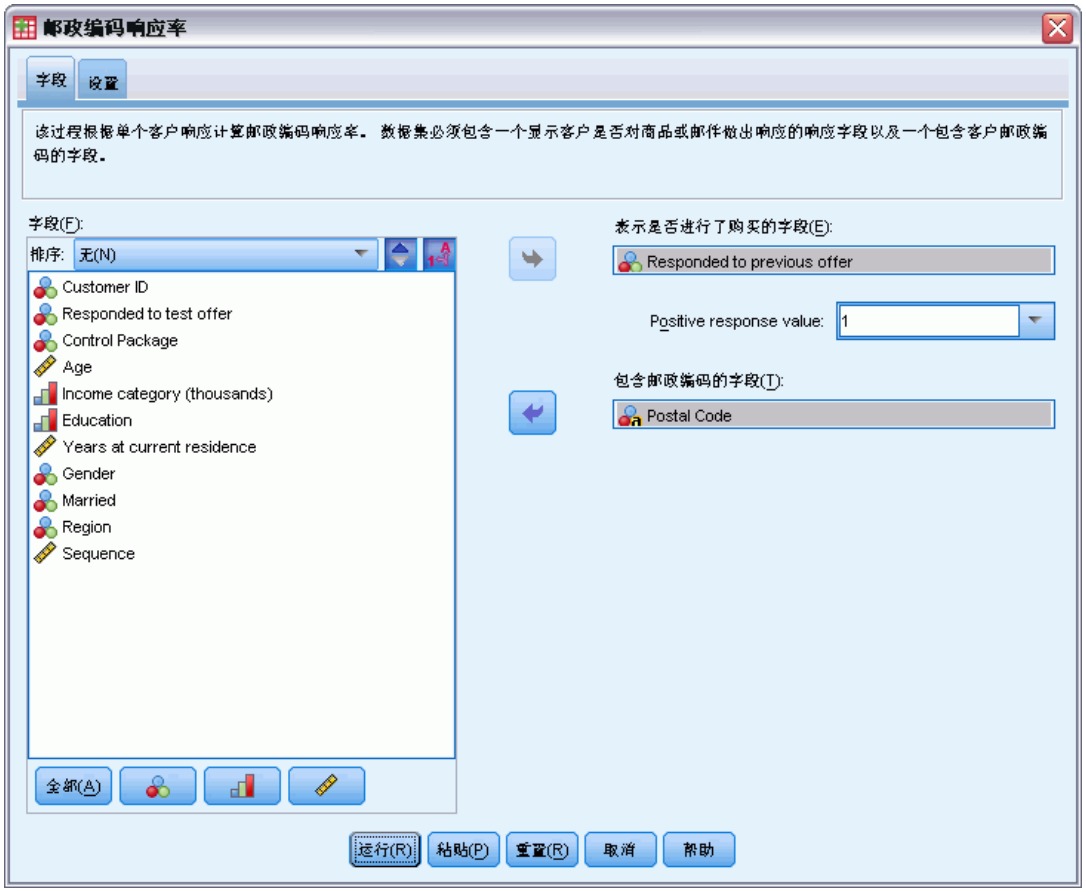
响应字段应当为分类字段，且具有一个代表所有正响应的值。任何其他非缺失值均被假设为负响应。如果响应字段代表连续（尺度）值，例如购买数量或购买金额，则需要创建新字段，以便为所有非零响应值指定单个正响应值。 有关详细信息，请参阅第 28 页码第 5 章中的创建分类响应字段。

运行分析

- 要计算“邮政编码响应率”，请从菜单中选择：
  直销 > **选择方法**

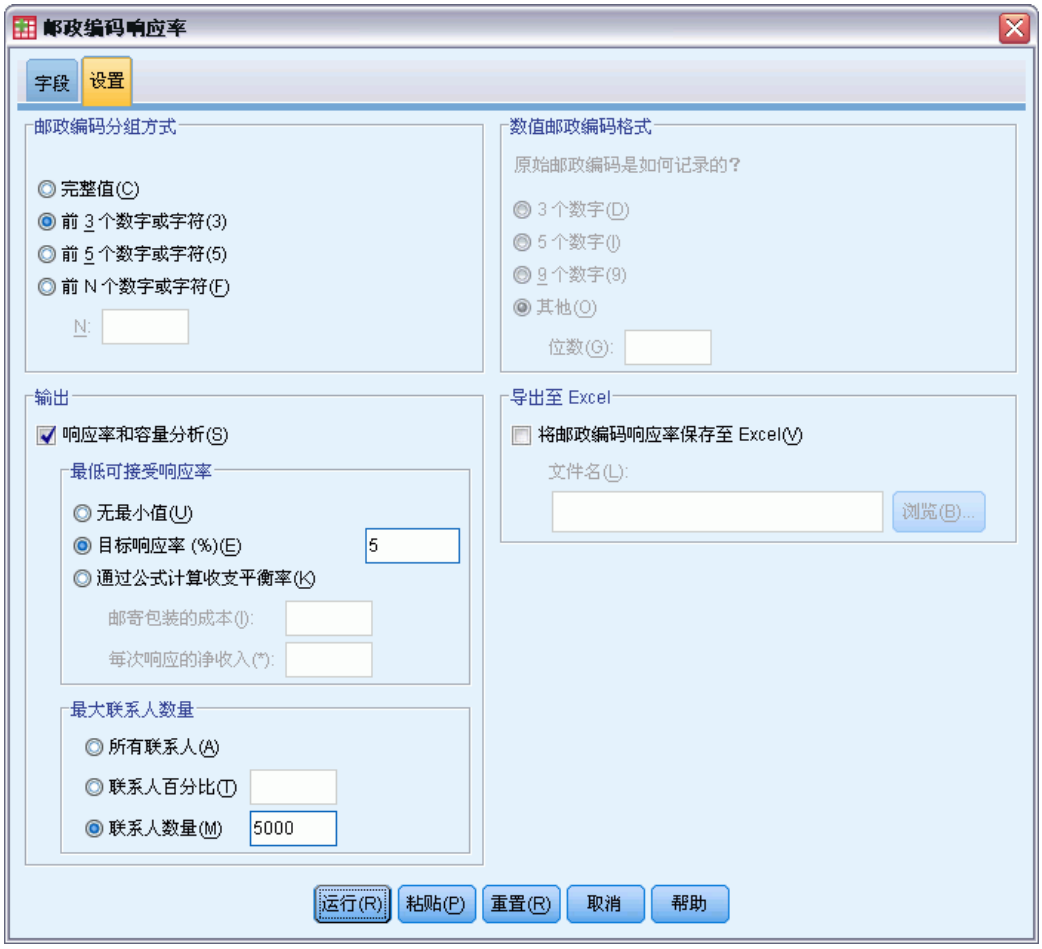
- 选择**标识最佳响应邮政编码**，然后单击**继续**。

图片 11-1
“邮政编码响应率：字段”选项卡



- 对于响应字段，选择已对先前产品做出响应。
- 对于正响应值，从下拉列表中选择是。在文本字段中显示值 1，因为“是”实际上为与记录值 1 关联的值标签。（如果正响应值未定义有值标签，您只需在文本字段中输入值即可。）
- 对于邮政编码字段，选择邮政编码。
- 单击**设置**选项卡。

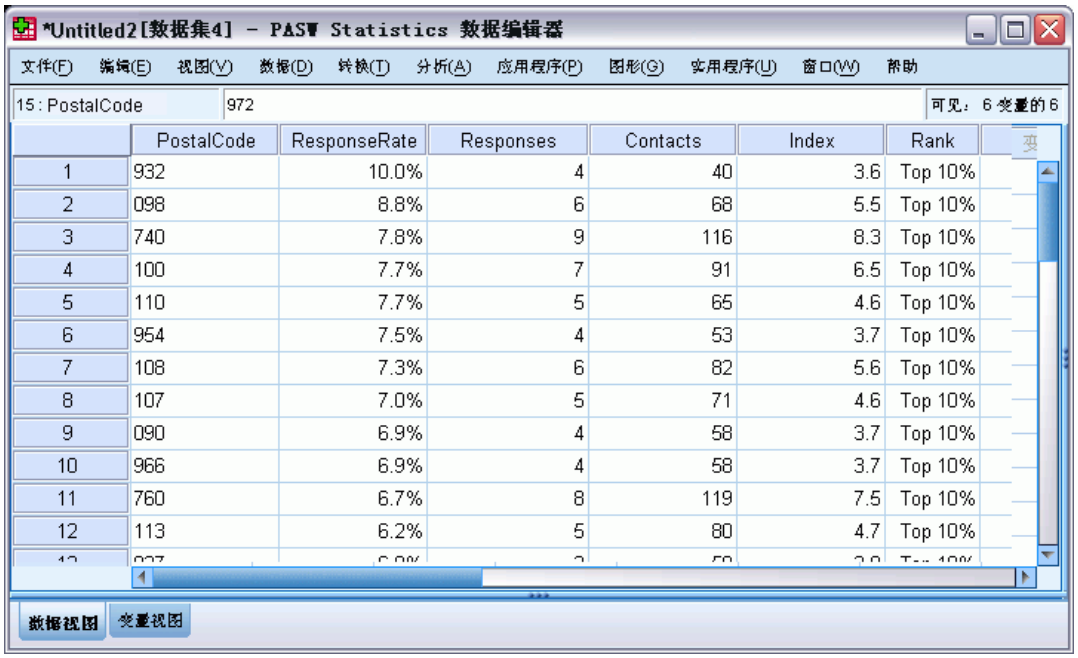
图片 11-2
“邮政编码响应率：设置”选项卡



- 在“邮政编码分组方式”组中，选择**前 3 个数字或字符**。这将计算所有具有以相同三个数字或字符开头邮政编码的联系人的组合响应率。例如，美国邮政编码的前三个数字代表大于全部 5 个数字邮政编码所定义地理区域的一般地理区域。
- 在“输出”组中，选择（选中）响应率和容量分析。
- 选择目标响应率并输入值 5。
- 选择联系人数量并输入值 5000。
- 然后单击**运行**以运行该过程。

输出

图片 11-3
包括邮政编码响应率的新数据集



	PostalCode	ResponseRate	Responses	Contacts	Index	Rank
1	932	10.0%	4	40	3.6	Top 10%
2	098	8.8%	6	68	5.5	Top 10%
3	740	7.8%	9	116	8.3	Top 10%
4	100	7.7%	7	91	6.5	Top 10%
5	110	7.7%	5	65	4.6	Top 10%
6	954	7.5%	4	53	3.7	Top 10%
7	108	7.3%	6	82	5.6	Top 10%
8	107	7.0%	5	71	4.6	Top 10%
9	090	6.9%	4	58	3.7	Top 10%
10	966	6.9%	4	58	3.7	Top 10%
11	760	6.7%	8	119	7.5	Top 10%
12	113	6.2%	5	80	4.7	Top 10%

这将自动创建新的数据集。此数据集为每个邮政编码包含单个记录（行）。在本例中，每行包含以相同的前三个数字或字符开头的所有邮政编码的摘要信息。

除包含邮政编码的字段外，新数据集还包含以下字段：

- **响应率。** 每个邮政编码中正响应的百分比。记录自动按响应率的降序进行排列；因此具有最高响应率的邮政编码出现在数据集的顶部。
- **响应。** 每个邮政编码中正响应的个数。
- **联系人。** 在每个邮政编码中包含响应字段的非缺失值的联系人总数。
- **指标。**基于公式 N x P x (1-P) 的“加权”响应，其中 N 为联系人数量，P 为以比例表示的响应率。对于具有相同响应率的两个邮政编码，该公式将为具有较多联系人的邮政编码分配更高的指标值。
- **排序。** 以降序排列的累积邮政编码响应率的十分位数排序（前 10%，前 20%，等等）。

由于我们在“邮政编码响应率”对话框的“设置”选项卡上选择了响应率和容量分析，因此在查看器中将显示摘要响应率表格和图表。

图片 11-4
响应率表格

Percentile	响应率	Contacts	Cumulative Response Rate	Total Contacts
Top 10%	7.3	1001	7.3	1001
Top 20%	5.3	956	6.3	1957
Top 30%	4.3	1042	5.6	2999
Top 40%	3.5	1127	5.1	4126
Top 50%	3.0	1173	4.6	5299
Top 60%	2.4	914	4.3	6213
Top 70%	2.0	948	4.0	7161
Top 80%	1.7	1095	3.7	8256
Top 90%	1.2	680	3.5	8936
Top 100%	.0	1064	3.1	10000

绿色: 满足目标响应率。
红色: 不满足目标响应率。

该表格按十分位数的降序排序（前 10%、前 20%，等等）汇总结果。

- 累积响应率是当前和所有前面行中正响应的组合百分比。由于结果以响应率的降序显示，因此这是当前十分位数和所有具有较高响应率的十分位数的组合响应率。
- 此表格基于您输入的目标响应率和最大联系人数量进行颜色编码。累积响应率等于或大于 5% 和累积联系人为 5,000 或更少的行显示为绿色。颜色编码基于首先达到的阈值。在本示例中，两个阈值在相同十分位数上达到。

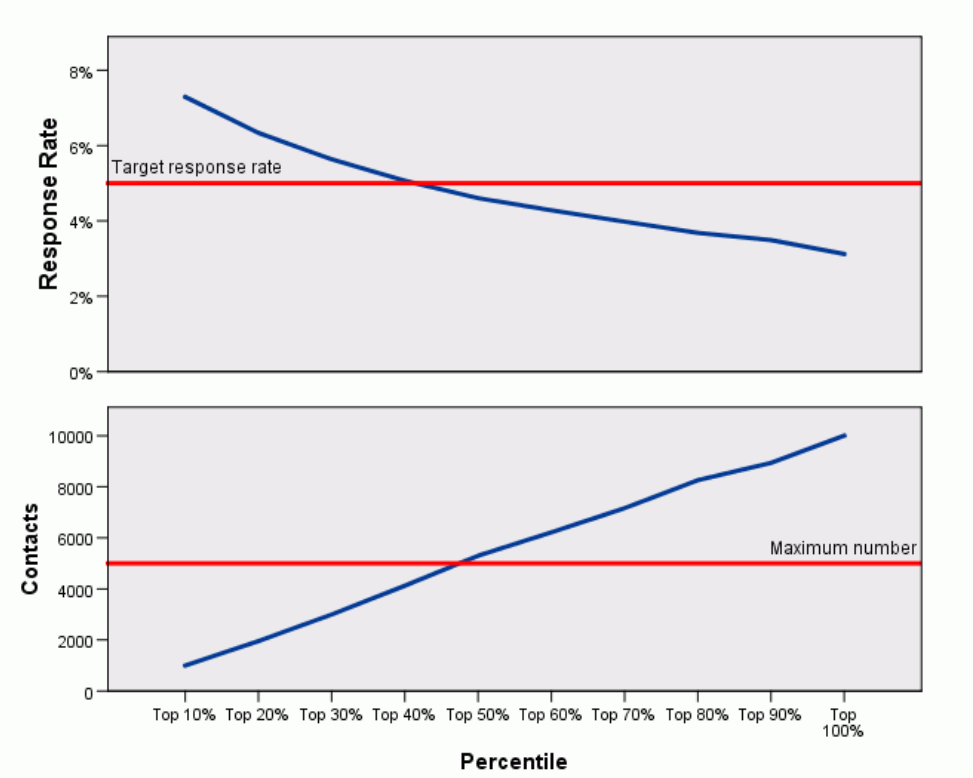
图片 11-5
响应率表格的智能输出

响应率表按十分位数的降序排序（前 10%、前 20%，等等）摘要列出结果。 累积响应率是当前和所有前面行中正响应的组合百分比。 由于结果以响应率的降序显示，因此这是当前十分位数和所有具有较高响应率的十分位数的组合响应率。 由于十分位数排序包含在新的数据集中，因此您可以方便地确定满足特定累积响应率要求的邮政编码。 在新的数据集中确定十分位数排序的字段称为"排序"，其中 1=前 10%、2=前 20% 等等。

指定的最小响应率为 5.00%。 指定的最大联系人数量为 5000。 表中的颜色编码基于首先达到的阈值。 两个阈值在相同类别上达到。 绿色行的累积响应率大于或等于指定的最小响应率，其累积联系人数量小于或等于指定的最大联系人数量。 红色行的累积响应率小于指定的最小响应率，其累积联系人数量大于指定的最大联系人数量。

该表带有关于如何阅读此表的一般说明文本。如果您指定了最小响应率或最大联系人数量，则它还包括有关结果与您指定阈值的关联程度的说明部分。

图片 11-6
累积响应率图表

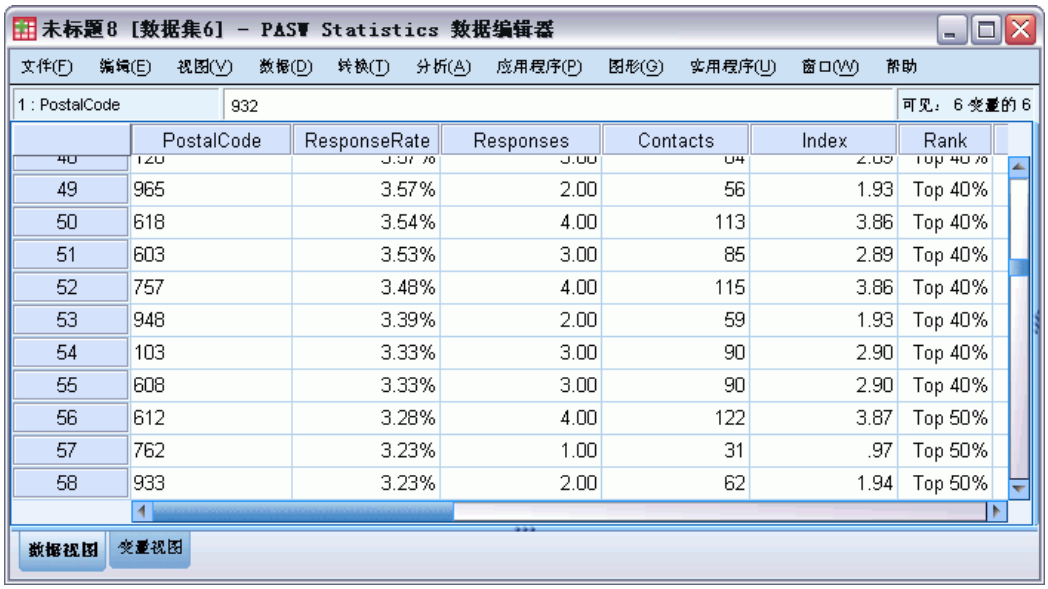


每个十分位数中累积响应率和累积联系人数量图表为在响应率表格中所显示的相同信息的可视化表示。最小累积响应率和最大累积联系人数量阈值在第 40 和 50 个百分位数之间的某个位置达到。

- 由于此图表按响应率十分位数的降序排序显示累积响应率，因此对于每个后续十分位数而言累积响应率行始终在下降。
- 由于联系人数量行代表累积联系人数量，因此它始终在上升。

通过表格和图表中的信息可知，如果要达到至少 5% 的响应率，但又不愿在活动中包含超过 5,000 个联系人，您应当重点关注前 4 个十分位数中的邮政编码。由于十分位数排序包含在新的数据集中，因此您可以方便地确定满足前 40% 要求的邮政编码。

图片 11-7
新数据集



	PostalCode	ResponseRate	Responses	Contacts	Index	Rank
49	965	3.57%	2.00	56	1.93	Top 40%
50	618	3.54%	4.00	113	3.86	Top 40%
51	603	3.53%	3.00	85	2.89	Top 40%
52	757	3.48%	4.00	115	3.86	Top 40%
53	948	3.39%	2.00	59	1.93	Top 40%
54	103	3.33%	3.00	90	2.90	Top 40%
55	608	3.33%	3.00	90	2.90	Top 40%
56	612	3.28%	4.00	122	3.87	Top 50%
57	762	3.23%	1.00	31	.97	Top 50%
58	933	3.23%	2.00	62	1.94	Top 50%

注意：排序记录为从 1 到 10 的整数值。此字段已定义有值标签，其中 1 = 前 10%，2 = 前 20%，依此类推。您将在数据编辑器的数据视图中看到实际排序值或值标签，具体取决于您的视图设置。

摘要

邮政编码响应率过程使用先前活动的结果来计算邮政编码响应率。这些响应率可以用于在未来的活动中集中面向特定的邮政编码。该过程创建一个包括每个邮政编码响应率的新数据集。根据响应率表格和图表中的信息，以及新数据集中的十分位数排序信息，您可以确定满足指定的最小累积响应率和/或最大累积联系人数量的一系列邮政编码。

章
12

购买倾向

购买倾向使用测试邮件或先前活动的结果来生成倾向得分。这些得分显示根据各种所选特征，哪些联系人最有可能做出响应。

此方法采用二元 Logistic 回归构建预测模型。构建并应用预测模型的过程包含两个基本步骤：

- 构建模型并保存模型文件。使用数据集构建兴趣结果（通常被称为**目标**）已知的模型。例如，如果您希望构建可预测谁可能会响应直接邮寄活动的模型，则需要从已包含响应人和未响应人信息的数据集开始。例如，这可能是对一小组客户发送的测试邮件的结果或来自过去类似活动的响应信息。
- 应用该模型到其他数据集（其中兴趣结果未知）以获取预测结果。

本示例使用两个数据文件：dmdata2.sav 用于构建模型，然后将模型应用到 dmdata3.sav。 有关详细信息，请参阅第 90 页码附录 A 中的样本文件。

数据注意事项

响应字段（兴趣目标结果）应当为分类字段，且具有一个代表所有正响应的值。任何其他非缺失值均被假设为负响应。如果响应字段代表连续（尺度）值，例如购买数量或购买金额，则需要创建新字段，以便为所有非零响应值指定单个正响应值。有关详细信息，请参阅第 35 页码第 6 章中的创建分类响应字段。

构建预测模型

- 打开数据文件 dmdata2.sav。

该文件包含收到测试邮件的人们的各种人口统计学特征，它还包含这些人是否响应邮件的信息。该信息记录在字段（变量）响应中。值为 1 表示联系人对邮件做出响应，值为 0 表示联系人未做出响应。

图片 12-1
数据编辑器中的数据文件内容

ID	做出响应	上一个	控制包装	邮政编码	年龄	收入	教育程度	家庭成员人数	性别
06262	0	0	1	96600	67	3	5	10	1
03692	0	0	0	95510	53	3	4	9	0
01480	0	0	1	92590	56	3	2	11	1
06118	0	0	1	92670	55	2	3	8	1
07378	0	0	1	92690	56	2	1	6	1
08467	0	0	0	93410	50	1	1	8	1
09621	0	0	1	93480	67	4	2	9	1
03029	0	0	0	93490	47	1	2	10	1
02660	0	0	1	93490	58	4	5	11	1

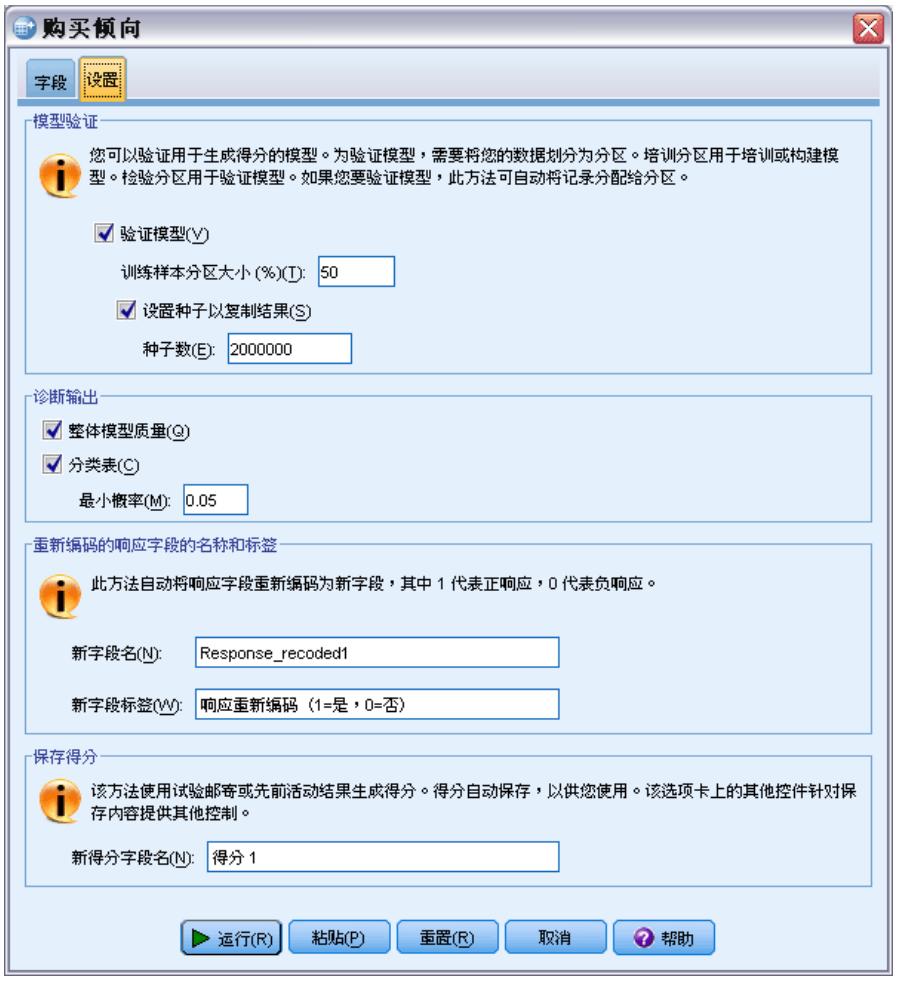
- 从菜单中选择：
  直销 > 选择方法

- 选择**选择最有可能购买的联系人**，并单击**继续**。

图片 12-2
“购买倾向：字段”选项卡

- 对于响应字段，选择已对测试产品做出响应。
- 对于正响应值，从下拉列表中选择是。在文本字段中显示值 1，因为“是”实际上为与记录值 1 关联的值标签。（如果正响应值未定义有值标签，您只需在文本字段中输入值即可。）
- 对于“预测倾向”，选择年龄、收入类别、教育、当前地址居住年限、性别、已婚、地区和孩子。
- 选择（勾选）**将模型信息导出到 XML 文件**。
- 单击**浏览**导航到要用于保存文件的位置，并输入文件名称。
- 在“购买倾向”对话框中，单击**设置**选项卡。

图片 12-3
“购买倾向：设置”选项卡



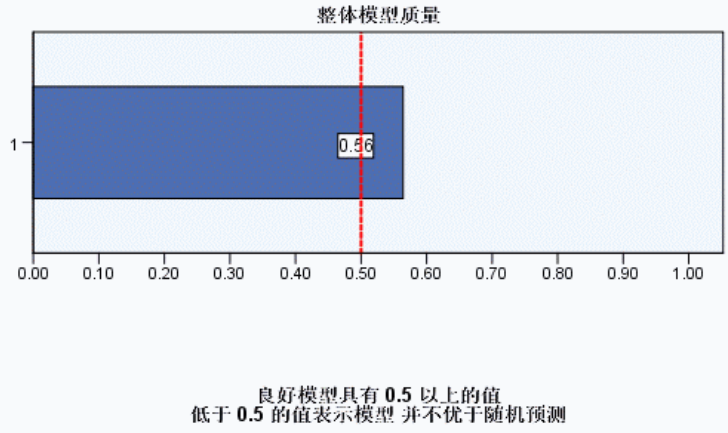
- 在模型验证组中，选择（选中）**验证模型**和**设置种子以复制结果**。
- 使用默认的训练样本分区大小 50% 和默认种子值 2000000。
- 在“诊断输出”组中，选择（选中）**整体模型质量**和**分类表**。
- 对于“最小可能性”，输入 0.05。作为一般规则，您必须指定一个接近最低目标响应率的比例值。值 0.05 表示响应率为 5%。
- 单击**运行**以运行过程并生成模型。

评估模型

“购买倾向”生成一个整体模型质量图表和分类表，以用于评估模型。

整体模型质量图表提供有关模型质量的快速直观指示。作为一般规则，整体模型质量应当大于 0.5。

图片 12-4
整体模型质量图表



要确认模型是否足以用于评分，您还应该检查分类表。

图片 12-5
分类表

分类表

已观测		已预测					
		训练样本			测试样本		
		响应重新编码（1=是，0=否）			响应重新编码（1=是，0=否）		
		否	是	百分比校正	否	是	百分比校正
响应重新编码（1=是，0=否）	否	651	249	72.33	653	267	70.98
	是	19	20	51.28	36	22	37.93
总计百分比		2.84	7.43	71.46	5.22	7.61	69.02

分类表将目标字段的预测值与目标字段的实际值相比较。总体准确率可以在某些方面显示模型工作情况，不过，如果构建模型是为了确定可能产生等于或大于指定的最小正响应率的正响应率的一组联系人，那么您可能更关注正确预测的正响应的百分比。

在本例中，分类表拆分为**训练样本**和**测试样本**。训练样本用于构建模型。然后，该模型被应用到测试样本，以查看模型的表现情况。

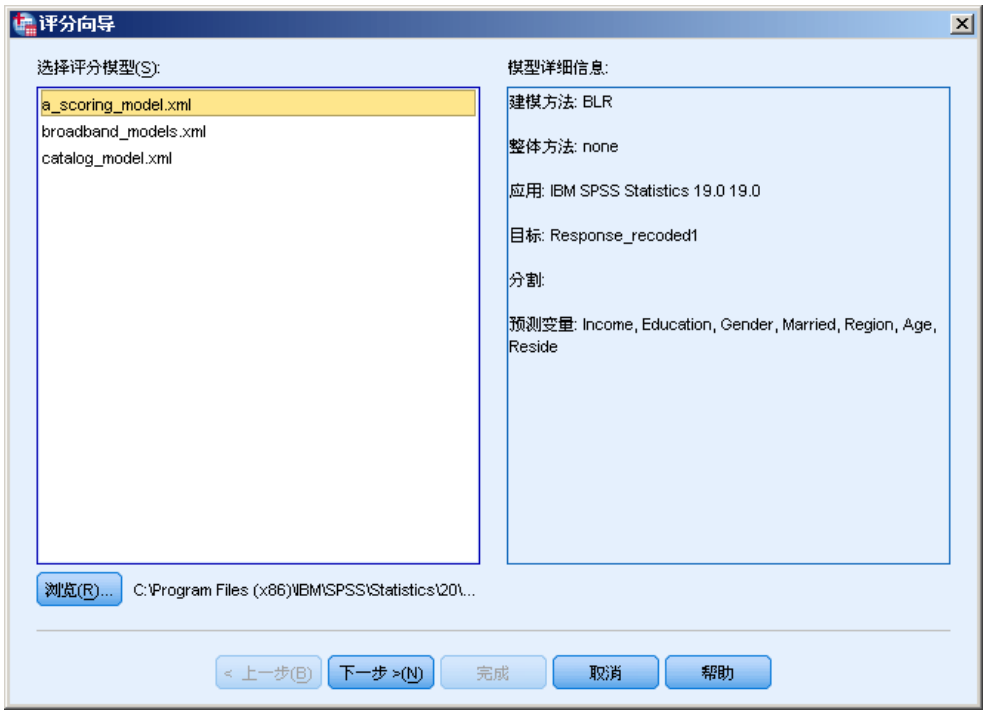
指定的最小响应率为 0.05 或 5%。分类表显示训练样本中正面响应者的正确分类率为 7.43%，而测试样本中为 7.61%。由于测试样本响应率大于 5%，因此该模型应当能够确定可能产生大于 5% 的响应率的一组联系人。

应用模型

- 打开数据文件 dmdata3.sav。该数据文件包含所有未在测试邮件中包含的联系人的人口统计学及其他信息。有关详细信息，请参阅第 90 页码附录 A 中的样本文件。

▶ 打开评分向导。要打开评分向导，从菜单中选择：
实用程序 > 评分向导

图片 12-6
评分向导，选择评分模型



▶ 单击**浏览**导航到要保存模型 XML 文件的位置，然后在浏览器对话框中单击**选择**。

在评分向导中将显示扩展名为 .xml 或 .zip 的所有文件。如果选定文件被识别为有效的模型文件，则显示模型的描述说明。

▶ 选择您创建的模型 XML 文件，然后单击**下一步**。

图片 12-7
评分向导，匹配模型字段

为了对活动数据集进行评分，该数据集必须包含对应于模型中的所有预测变量的字段（变量）。如果模型还包含拆分字段，那么该数据集还必须包含对应于模型中所有拆分字段的字段。

- 默认情况下，自动匹配活动数据集中任何与模型中的字段具有相同名称和类型的字段。
- 使用下拉列表匹配数据集字段到模型字段。模型和数据集中每个字段的数据类型必须相同才能匹配字段。
- 在模型中的所有预测变量（以及拆分字段，如果有的话）与活动数据集中的字段匹配之前，您无法继续向导或对活动数据集进行评分。

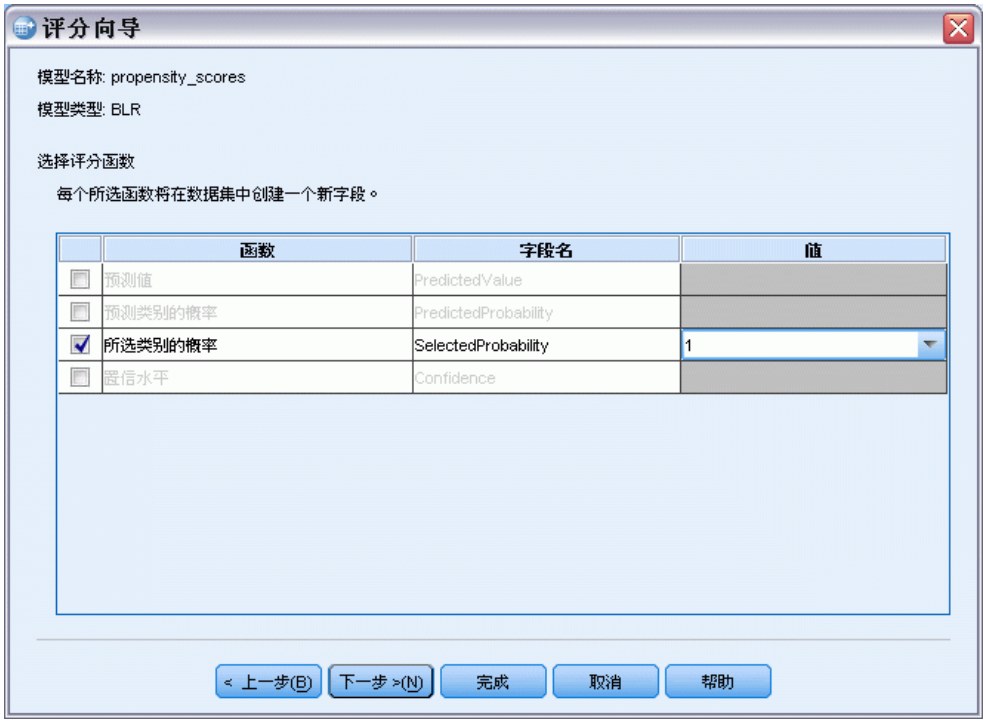
活动数据集不包含名为 Income 的字段。因此，与模型字段 Income 对应的“数据集字段”列中的单元格初始情况下为空白。您需要在活动数据集中选择等效于该模型字段的字段。

▶ 从 Income 模型字段对应行空白单元格的“数据集字段”列的下拉列表中，选择 IncomeCategory。

注意： 除了字段名和类型以外，您需要确保要评分的数据集中的实际数据值的记录方式与构建模型的数据集中的数据值相同。例如，如果模型使用 Income 字段构建，后者将收入划分为四种类别，而活动数据集中的 IncomeCategory 则将收入划分为六种类别或四种不同的类别，因此这些字段实际上彼此并不匹配，结果得分将不可靠。

单击**下一步**继续执行评分向导的下一步。

图片 12-8
评分向导：选择评分函数



评分函数是所选模型可用的“得分”类型。可用的评分函数取决于模型。对于在本示例中使用的二项 logistic 模型，可用函数为预测值、预测值的概率、所选值的概率和置信度。

在本示例中，我们对邮件正响应的预测概率感兴趣，因此我们需要选定值的概率。

- 选择（选中）**选定类别的概率**。
- 在“值”列中，从下拉列表中选择 1。目标的可能值列表在模型中定义，并基于用于构建模型的数据文件中的目标值。
- 取消选择（取消选中）所有其他评分函数。
- 根据需要，可以为在活动数据集中包含得分值的新字段指定一个更具描述性的名称。例如，Probability_of_responding。
- 单击**完成**将模型应用到活动数据集。

包含正响应概率的新字段被附加到数据集的末尾。

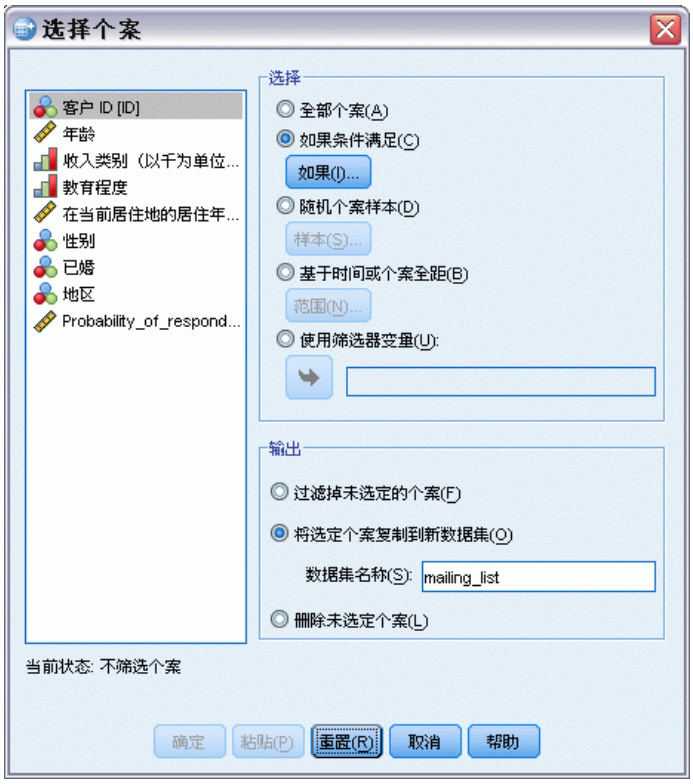
图片 12-9
包含新概率字段的数据集

家庭成员人数	性别	已婚	地区	Probability_of_responding
7	1	0	4	.04
9	0	0	4	.03
12	0	0	4	.03
8	0	0	4	.04
13	0	0	4	.07
10	0	0	4	.04
12	0	0	4	.03
15	0	0	4	.05
10	0	0	4	.05
14	0	0	4	.02
5	0	0	4	.12

然后，您可以使用该字段来选择可能产生等于或大于特定级别的正响应率的联系人子集。例如，可以创建包含可能产生至少 5% 正响应率的个案子集的新数据集。

▶ 从菜单中选择：
数据 > 选择个案

图片 12-10
“选择个案”对话框



▶ 在“选择个案”对话框中，选择**如果条件满足**，然后单击**如果**。

图片 12-11
选择个案:“如果”对话框

▶ 在“选择个案:如果”对话框中输入以下表达式：

```
Probability_of_responding >=.05
```

注意： 如果您为包含概率值的字段使用了不同的名称，请输入此名称，而不是 Probability_of_responding。默认名称为 SelectedProbability。

▶ 单击**继续**。

▶ 在“选择个案”对话框中，选择**将选定个案复制到新数据集**，然后输入新数据集的名称。数据集名称必须符合字段（变量）命名规则。

▶ 单击**确定**以使用选定联系人创建数据集。

新数据集只包含那些具有至少 5% 正响应的预测概率的联系人。

图片 12-12
带选定联系人的新数据集

家庭成员人数	性别	已婚	地区	Probability_of_responding
13	0	0	4	.07
15	0	0	4	.05
10	0	0	4	.05
5	0	0	4	.12
7	0	0	4	.08
10	0	0	4	.10
15	1	0	4	.05
11	1	0	4	.08
9	1	0	4	.08
9	1	0	4	.05

摘要

购买倾向使用测试邮件或先前活动的结果来生成倾向得分。这些得分显示根据各种所选特征，哪些联系人最有可能做出响应。此技术构建了一个可应用到数据集以获得倾向得分的预测模型。

章
13

控制包装检验

该方法比较市场营销活动，以检查不同包装或商品之间是否存在显著的效果差异。活动效果通过响应来测量。

例如，公司直销部门想了解新的包装设计能否产生比现有包装更多的正面响应。因此他们发出测试邮件，以确定新包装能否产生明显更高的正响应率。测试邮件包括获得现有包装的控制组和获得新包装设计的测试组。然后比较两组的结果，看看是否存在显著差异。

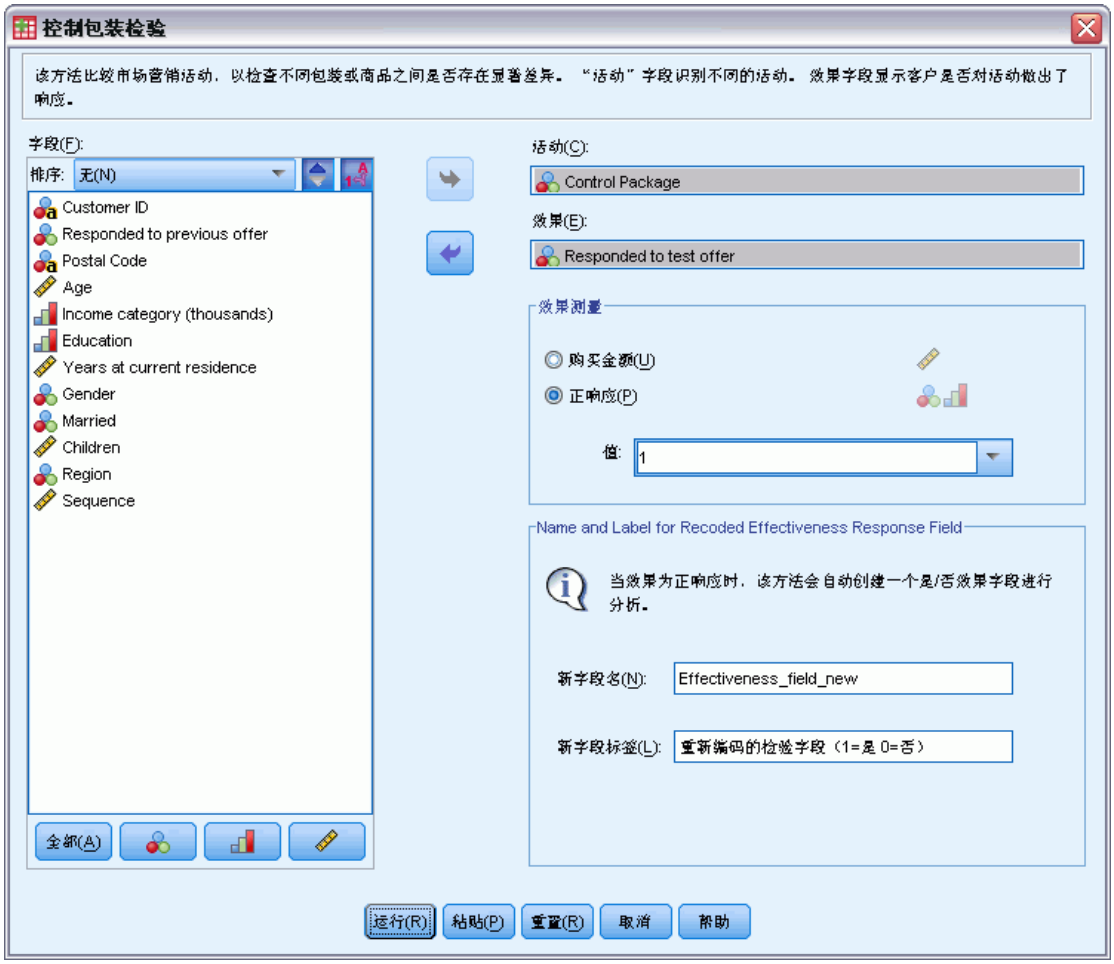
这些信息收集在 dmdata.sav 中。 有关详细信息，请参阅第 90 页码附录 A 中的样本文件。

运行分析

- 要获取控制包装检验，请从菜单中选择：
  直销 > 选择方法

- 选择**比较活动效果（控制包装检验）**并单击**继续**。

图片 13-1
“控制包装检验：字段”选项卡



- 对于“活动”字段，选择控制包装。
- 对于“效果响应”字段，选择已对测试产品做出响应。
- 选择“回应”。
- 对于正响应值，从下拉列表中选择是。在文本字段中显示值 1，因为“是”实际上为与记录值 1 关联的值标签。（如果正响应值未定义有值标签，您只需在文本字段中输入值即可。）

  将自动创建新字段，其中 1 代表正响应，0 代表负响应，并在此新字段上执行分析。您可以使用自己的名称和标签覆盖默认名称和标签。对于本示例，我们将使用已提供的字段名称。

- 单击运行以运行该过程。

输出

图片 13-2
控制包装检验输出

		控制包装			
		控制		检验	
		计数	列 N %	计数	列 N %
效果（1=是 0=否）	0	879	96.6%	984	97.7%
	1	31	3.4%	23	2.3%

在 控制 和 检验 之间不存在统计显著性差异。

过程输出包含两个表格，其中一个显示由“活动”字段定义的每个组的正、负响应计数与百分比，另一个则表明组响应率之间是否存在明显差异。

- 效果是响应字段的重新编码版本，其中 1 代表正响应，0 代表负响应。
- 控制包装的正响应率为 3.8%，而检验包装的正响应率为 6.2%。

表格下方的简单文本说明显示两组之间的差异非常显著，这表明检验包装的较高响应率可能不是随机概率的结果。该文本表将包含在分析中包括的每个可能成对组的比较。由于该示例中只有两组，因此只比较一次。如果超过 5 组，则使用“列比例比较”表来替换文本描述表。

摘要

“控制包装检验”比较市场营销活动，以检查不同包装或商品之间是否存在显著的效果差异。本例中，检验包装 6.2% 的正响应率明显高于控制包装 3.8% 的正响应率。这表明，您应该用新的包装设计替代旧的包装设计，但您还需考虑其他一些因素，比如与新的包装设计相关的任何附加成本。

附录
A

样本文件

随产品一起安装的样本文件可以在安装目录的 Samples 子目录中找到。对于以下每种语言在“样本”子目录中有单独的文件夹：英语、法语、德语、意大利语、日语、韩语、波兰语、俄语、简体中文、西班牙语和繁体中文。

并非所有样本文件均提供此处的全部语言版本。如果样本文件未提供某种语言的版本，则相应语言文件夹中包含该样本文件的英语版本。

描述

以下是对在整个文档的各种示例中使用的样本文件的简要描述。

- **accidents.sav。** 该假设数据文件涉及某保险公司，该公司正在研究给定区域内汽车事故的年龄和性别风险因子。每个个案对应一个年龄类别和性别类别的交叉分类。
- **adl.sav。**该假设数据文件涉及在确定针对脑卒中患者的建议治疗类型的优点方面的举措。医师将女性脑卒中患者随机分配到两组中的一组。第一组患者接受标准的物理治疗，而第二组患者则接受附加的情绪治疗。在进行治疗的三个月时间里，将为每个患者进行一般日常生活行为的能力评分并作为原始变量。
- **advert.sav。**该假设数据文件涉及某零售商在检查广告支出与销售业绩之间的关系方面的举措。为此，他们收集了过去的销售数据以及相关的广告成本..
- **aflatoxin.sav。**该假设数据文件涉及对谷物的黄曲霉毒素的检测，该毒素的浓度会因谷物产量的不同（不同谷物之间及同种谷物之间）而有较大变化。谷物加工机从 8 个谷物产量的每一个中收到 16 个样本并以十亿分之几 (PPB) 为单位来测量黄曲霉毒素的水平。
- **anorectic.sav。**在研究厌食/暴食行为的标准症状参照时，研究人员对 55 名已知存在进食障碍的青少年进行了调查。其中每名患者每年都将进行四次检查，因此总观测数为 220。在每次观测期间，将对这些患者按 16 种症状逐项评分。但 71 号和 76 号患者的症状得分均在时间点 2 缺失，47 号患者的症状得分在时间点 3 缺失，因此有效观测数为 217。
- **bankloan.sav。**该假设数据文件涉及某银行在降低贷款拖欠率方面的举措。该文件包含 850 位过去和潜在客户的财务和人口统计信息。前 700 个个案是以前曾获得贷款的客户。剩下的 150 个个案是潜在客户，银行需要按高或低信用风险对他们进行分类。
- **bankloan_binning.sav。**该假设数据文件包含 5,000 位过去客户的财务和人口统计信息。
- **behavior.sav。**在一个经典示例中，52 名学生被要求以 10 分的标度对 15 种情况和 15 种行为的组合进行评价，该 10 分的标度介于 0 =平均值在个人值之上，值被视为相异性。
- **behavior_ini.sav。**该数据文件包含 behavior.sav 的二维解的初始配置。

- **brakes.sav。**该假设数据文件涉及某生产高性能汽车盘式制动器的工厂的质量控制。该数据文件包含对 8 台专用机床中每一台的 16 个盘式制动器的直径测量。盘式制动器的目标直径为 322 毫米。
- **breakfast.sav。**在一项经典研究中，21 名 Wharton School MBA 学生及其配偶被要求按照喜好程度顺序对 15 种早餐食品进行评价，从 1 =他们的喜好根据六种不同的情况加以记录，从“全部喜欢”到“只带饮料的快餐”。
- **breakfast-overall.sav。**该数据文件只包含早餐食品喜好的第一种情况，即“全部喜欢”。
- **broadband_1.sav。**该假设数据文件包含各地区订制了全国宽带服务的客户的数量。该数据文件包含 4 年期间 85 个地区每月的订户数量。
- **broadband_2.sav。**该数据文件和 broadband_1.sav 一样，但包含另外三个月的数据。
- **car_insurance_claims.sav。**在别处被提出和分析的关于汽车损坏赔偿的数据集。平均理赔金额可以当作其具有 gamma 分布来建模，通过使用逆联接函数将因变量的均值与投保者年龄、车辆类型和车龄的线性组合关联。提出理赔的数量可以作为尺度权重。
- **car_sales.sav。**该数据文件包含假设销售估计值、订价以及各种品牌和型号的车辆的物理规格。订价和物理规格可以从 edmunds.com 和制造商处获得。
- **car_sales_uprepared.sav。**这是 car_sales.sav 的修改版本，不包含字段的任何已转换版本。
- **carpet.sav。**在一个常用示例中，一家公司非常重视一种新型地毯清洁用品的市场营销，希望检验以下五种因素对消费者偏好的影响—包装设计、品牌名称、价格、优秀家用品 标志和退货保证。包装设计有三个因子水平，每个因子水平因刷体位置而不同；有三个品牌名称（K2R、Glory 和 Bissell）；有三个价格水平；最后两个因素各有两个级别（有或无）。十名消费者对这些因素所定义的 22 个特征进行了排序。变量优选包含对每个特征的平均等级的排序。低排序与高偏好相对应。此变量反映了对每个特征的偏好的总体度量。
- **carpet_prefs.sav。**该数据文件所基于的示例和在 carpet.sav 中所描述的一样，但它还包含从 10 位消费者的每一位中收集到的实际排列顺序。消费者被要求按照从最喜欢到最不喜欢的顺序对 22 个产品特征进行排序。carpet_plan.sav 中定义了变量 PREF1 到 PREF22 包含相关特征的标识符。
- **catalog.sav。**该数据文件包含某编目公司出售的三种产品的假设每月销售数据。同时还包括 5 个可能的预测变量的数据。
- **catalog_seasfac.sav。**除添加了一组从“季节性分解”过程中计算出来的季节性因子和附带的日期变量外，该数据文件和 catalog.sav 是相同的。
- **cellular.sav。**该假设数据文件涉及某便携式电话公司在减少客户流失方面的举措。客户流失倾向分被应用到帐户，分数范围从 0 到 100。得到 50 分或更高分数的帐户可能会更换提供商。
- **ceramics.sav。**该假设数据文件涉及某制造商在确定新型优质合金是否比标准合金具有更高的耐热性方面的举措。每个个案代表对一种合金的单独检验；个案中会记录合金的耐热极限。
- **cereal.sav。**该假设数据文件涉及一份 880 人参于的关于早餐喜好的民意调查，该调查记录了参与者的年龄、性别、婚姻状况以及生活方式是否积极（根据他们是否每周至少做两次运动）。每个个案代表一个单独的调查对象。

- **clothing_defects.sav。**这是关于某服装厂的质量控制过程的假设数据文件。检验员要对工厂中每次大批量生产的服装进行抽样检测并清点不合格的服装的数量。
- **coffee.sav。**这是关于六种冰咖啡的认知品牌形象的数据文件。对于 23 种冰咖啡特征属性中的每种属性，人们选择了由该属性所描述的所有品牌。为保密起见，六种品牌用 AA、BB、CC、DD、EE 和 FF 来表示。
- **contacts.sav。**该假设数据文件涉及一组公司计算机销售代表的联系方式列表。根据这些销售代表所在的公司部门及其公司的秩来对每个联系方式进行分类。同时还记录了最近一次的销售量、最近一次销售距今的时间和所联系公司的规模。
- **creditpromo.sav。**该假设数据文件涉及某百货公司在评价最新信用卡促销的效果方面的举措。为此，随机选择了 500 位持卡人。其中一半收到了宣传关于在接下来的三个月内降低消费利率的广告。另一半收到了标准的季节性广告。
- **customer_dbase.sav。** 该假设数据文件涉及某公司在使用数据仓库中的信息来为最有可能回应的客户提供特惠商品方面的举措。随机选择客户群的子集并为其提供特惠商品，同时记录下他们的回应。
- **customer_information.sav。**该假设数据文件包含客户邮寄信息，如姓名和地址。
- **customer_subset.sav。**来自 customer_dbase.sav 的拥有 80 个个案的子集。
- **debate.sav。**该假设数据文件涉及在某政治辩论前后对该辩论的参与者所做的调查的成对回答。每个个案对应一个单独的调查对象。
- **debate_aggregate.sav。**该假设数据文件分类汇总了 debate.sav 中的回答。每个个案对应一个辩论前后的偏好的交叉分类。
- **demo.sav。**这是关于购物客户数据库的假设数据文件，用于寄出每月的商品。将记录客户对商品是否有回应以及各种人口统计信息。
- **demo_cs_1.sav。**该假设数据文件涉及某公司在汇编调查信息数据库方面的举措的第一步。每个个案对应不同的城市，并记录地区、省、区和城市标识。
- **demo_cs_2.sav。**该假设数据文件涉及某公司在汇编调查信息数据库方面的举措的第二步。每个个案对应来自第一步中所选城市的不同的家庭单元，并记录地区、省、区、市、子区和单元标识。还包括设计前两个阶段的抽样信息。
- **demo_cs.sav。**该假设数据文件包含用复杂抽样设计收集的调查信息。每个个案对应不同的家庭单元，并记录各种人口统计和抽样信息。
- **dmdata.sav。**该假设数据文件包含直销公司的人口统计学和购买信息。dmdata2.sav 包含收到测试邮件的联系人子集的信息，dmdata3.sav 包含未收到测试邮件的其余联系人的信息。
- **dietstudy.sav。**该假设数据文件包含对 “Stillman diet” 的研究结果。每个个案对应一个单独的主体，并记录其在实行饮食方案前后的体重（磅）以及甘油三酸酯的水平（毫克/100 毫升）。
- **dvdplayer.sav。**这是关于开发新的 DVD 播放器的假设数据文件。营销团队用原型收集了焦点小组数据。每个个案对应一个单独的被调查用户，并记录他们的人口统计信息及其对原型问题的回答。
- **german_credit.sav。**该数据文件取自加州大学欧文分校的 Repository of Machine Learning Databases 中的 “German credit” 数据集。
- **grocery_1month.sav。**该假设数据文件是在数据文件 grocery_coupons.sav 的基础上加上了每周购物“累计”，所以每个个案对应一个单独的客户。所以，一些每周更改的变量消失了，而且现在记录的消费金额是为期四周的研究过程中的消费金额之和。

- **grocery_coupons.sav。**该假设数据文件包含由重视顾客购物习惯的杂货连锁店收集的调查数据。对每位顾客调查四周，每个个案对应一个单独的顾客周，并记录有关顾客购物地点和方式的信息（包括那一周里顾客在杂货上的消费金额）。
- **guttman.sav。** Bell 创建了一个表，用来阐释可能的社会群体。Guttman 引用了该表的一部分，其中包括五个变量，用于描述以下七个理论社会群体的社会交往、对群体的归属感、成员的物理亲近度以及关系正式性：观众（比如在足球比赛现场的人们）、听众（比如在剧院或听课堂讲座的人们）、公众（比如报纸或电视观众）、组织群体（与观众类似但具有紧密的关系）、初级群体（关系密切）、次级群体（自发组织）及现代社区（因在物理上亲近而导致关系松散并需要专业化服务）。
- **health_funding.sav。**该假设数据文件包含关于保健基金（每 100 人的金额）、发病率（每 10,000 人的比率）以及保健提供商拜访率（每 10,000 的比率）的数据。每个个案代表不同的城市。
- **hivassay.sav。**该假设数据文件涉及某药物实验室在开发用于检测 HIV 感染的快速化验方面的举措。化验结果为八个加深的红色阴影，如果有更深的阴影则表示感染的可能性很大。用 2,000 份血液样本来进行实验室试验，其中一半受到 HIV 感染而另一半没有受到感染。
- **hourlywagedata.sav。**该假设数据文件涉及在政府机关和医院工作的具有不同经验水平的护士的时薪。
- **insurance_claims.sav。**该假设数据文件涉及某保险公司，该公司希望构建一个模型用于标记可疑的、具有潜在欺骗性的理赔。每个个案代表一次单独的理赔。
- **insure.sav。**该假设数据文件涉及某保险公司，该公司正在研究指示客户是否会根据 10 年的人寿保险合同提出理赔的风险因子。数据文件中的每个个案代表一副根据年龄和性别进行匹配的合同，其中一份记录了一次理赔而另一份则没有。
- **judges.sav。**该假设数据文件涉及经过训练的裁判（加上一个体操爱好者）对 300 次体操表演给出的分数。每行代表一次单独的表演；裁判们观看相同的表演。
- **kinship_dat.sav。**Rosenberg 和 Kim 开始分析 15 个亲属关系项（伯母、兄弟、表兄妹、女儿、父亲、孙女、祖父、祖母、孙子、母亲、侄子或外甥、侄女或外甥女、姐妹、儿子和叔叔）。他们让四组大学生（两组女同学，两组男同学）根据相似程度将各项排序。他们让其中的两组同学（一组女同学，一组男同学）进行了两次排序，第二次排序和第一次排序采取的标准不同。这样，一共得到六组“源”。每个源对应一个 15×15 的相似性矩阵，其单元格中的值等于源中的人数减去此源中对象被划分的次数。
- **kinship_ini.sav。**该数据文件包含 kinship_dat.sav 的三维解的初始配置。
- **kinship_var.sav。**该数据文件包含自变量 gender、gener(ation) 和 degree (of separation)，这些变量可用于解释 kinship_dat.sav 的解的维数。具体而言，它们可用来将解的空间限制为这些变量的线性组合。
- **marketvalues.sav。**该数据文件涉及 1999 - 2000 年间 Algonquin，Ill. 地区新的房屋开发中的住房销售。这些销售仅仅来自公众记录。
- **nhis2000_subset.sav。**美国健康访问调查 (NHIS) 是针对美国全体公民的大型人口调查。该调查对美国的具有全国代表性的家庭样本进行了面对面的访问，并获取了每个家庭的成员的健康行为和健康状态的人口统计信息和观察数据。该数据文件包含取自 2000 年调查信息的子集。国家健康统计中心。2000 年美国健康访问调查。公用数据文件和文档。

ftp://ftp.cdc.gov/pub/Health_Statistics/NCHS/Datasets/NHIS/2000/。2003 年发布。

- **ozone.sav。**这些数据包含了用来根据其余变量预测臭氧浓度的六个气象变量的 330 个观察值。在以前的研究人员中， 和 发现了这些变量之间的非线性，这妨碍了标准回归方法。
- **pain_medication.sav。**该假设数据文件包含用于治疗慢性关节炎疼痛的抗炎药的临床试验结果。我们感兴趣的是该药见效的时间以及它和现有药物的比较。
- **patient_los.sav。**该假设数据文件包含被医院确诊为疑似心肌梗塞（即 MI 或“心脏病发作”）的患者的治疗记录。每个个案对应一位单独的患者，并记录与其住院期有关的一些变量。
- **patlos_sample.sav。**该假设数据文件包含在治疗心肌梗塞（即 MI 或“心脏病发作”）期间收到溶解血栓剂的患者的样本治疗记录。每个个案对应一位单独的患者，并记录与其住院期有关的一些变量。
- **poll_cs.sav。**该假设数据文件涉及民意测验专家在确定正式立法前公众对法案的支持水平方面的举措。个案对应注册的选民。每个个案记录选民居住的县、镇、区。
- **poll_cs_sample.sav。**该假设数据文件包含在 poll_cs.sav 中列出的选民的样本。该样本是根据 poll.csplan 中指定的设计来选取的，而且该数据文件记录包含概率和样本权重。请注意，由于该抽样计划使用与大小成正比 (PPS) 方法，因此，还有一个文件 (poll_jointprob.sav) 包含联合选择概率。在选取了样本之后，对应于选民人群统计信息及其对提交法案的意见的附加变量将被收集并添加到数据文件。
- **property_assess.sav。**该假设数据文件涉及某县资产评估员在利用有限的资源不断更新资产价值评估方面的举措。个案对应过去一年中县里所出售的资产。数据文件中的每个个案记录资产所在的镇、最后评估资产的评估员、该次评估距今的时间、当时的估价以及资产的出售价格。
- **property_assess_cs.sav。**该假设数据文件涉及某州资产评估员在利用有限的资源不断更新资产价值评估方面的举措。个案对应该州的资产。数据文件中的每个个案记录资产所在的县、镇和区，最后一次评估距今的时间以及当时的估价。
- **property_assess_cs_sample.sav。**该假设数据文件包含在 property_assess_cs.sav 中列出的资产的样本。该样本是根据 property_assess.csplan 中指定的设计来选取的，而且该数据文件记录包含概率和样本权重。在选取了样本之后，附加变量 Current value 将被收集并添加到数据文件。
- **recidivism.sav。**该假设数据文件涉及某政府执法机构在了解其管辖区域内的屡犯率方面的举措。每个个案对应先前的一名罪犯，并记录其人口统计信息和第一次犯罪的详细资料；如果在第一次被捕后两年内又第二次被捕，则还将记录两次被捕间隔的时间。
- **recidivism_cs_sample.sav。**该假设数据文件涉及某政府执法机构在了解其管辖区域内的屡犯率方面的举措。每个个案对应在 2003 年 6 月期间第一次被捕释放的先前的一名罪犯，并记录其人口统计信息和第一次犯罪的详细资料，及其第二次被捕的数据（如果发生在 2006 年 6 月底之前）。根据 recidivism_cs.csplan 中指定的抽样计划从抽样部门选择罪犯；该计划使用与大小成正比 (PPS) 方法，因此，还有一个文件 (recidivism_cs_jointprob.sav) 包含联合选择概率。
- **rfm_transactions.sav。** 此假设数据文件包含购买交易数据，即每笔交易的购买日期、购买商品和消费金额。

- **salesperformance.sav。**这是关于评估两个新的销售培训课程的假设数据文件。60 名员工被分成 3 组且都接受标准的培训。另外，组 2 接受技术培训；组 3 接受实践教程。在培训课程结束时，对每名员工进行测验并记录他们的分数。数据文件中的每个个案代表一名单独的受训者，并记录其被分配到的组以及测验的分数。
- **satisf.sav。**该假设数据文件涉及某零售公司在 4 个商店位置所进行的满意度调查。总共对 582 位客户进行了调查，每个个案代表一位单独客户的回答。
- **screws.sav。**该数据文件包含关于螺钉、螺栓、螺母和图钉的特征的信息。
- **shampoo_ph.sav**这是关于某发制品厂的质量控制的假设数据文件。在规定的时间间隔对六批独立输出的产品进行检测并记录它们的 pH 值。目标范围是 4.5 - 5.5。
- **ships.sav。**在别处被提出和分析的关于波浪对货船造成的损坏的数据集。在给定了船的类型、建造工期和服务期后，可以根据以泊松比率发生来为事件计数建模。在因子交叉分类构成的表格中，每个单元格的分类汇总服务月数提供遇到风险的值。
- **site.sav。**该假设数据文件涉及某公司在为扩展业务而选择新址方面的举措。该公司聘请了两名顾问分别对选址进行评估，除了提供长期报告外，他们还要以“前景颇佳”、“前景良好”或“前景不佳”来对每个选址进行总结。
- **smokers.sav。**该数据文件摘自 1998 年全国家庭药物滥用调查并且是美国家庭的概率样本。(http://dx.doi.org/10.3886/ICPSR02934) 因此，分析该数据文件的第一步应该是对数据进行加权以反映总体趋势。
- **stocks.sav** 该假设数据文件包含某一年的股票价格和成交量。
- **stroke_clean.sav。**该假设数据文件包含某医学数据库在经过“数据准备”选项中的过程清理后的状态。
- **stroke_invalid.sav。**该假设数据文件包含某医学数据库的初始状态及一些数据输入错误。
- **stroke_survival。**此假设数据文件涉及正在研究结束缺血性中风后复元计划的患者存活时间的研究人员面临着很多挑战。中风后，记录心肌梗塞、缺血性中风或出血性中风的发生及其时间。样本为左侧截短，因为只包含在中风后管理的复元计划结束后存活的患者。
- **stroke_valid.sav。**该假设数据文件包含在使用“验证数据”过程检查值后，某医学数据库的状态。它仍包含潜在异常个案。
- **survey_sample.sav。**此数据文件包含调查数据，包括人口统计学数据和各种态度测量。它基于 1998 NORC 综合社会调查的变量子集，但某些数据值已经过修改，并添加了其他虚拟变量以供演示用途。
- **telco.sav。**该假设数据文件涉及某电信公司在减少客户群中的客户流失方面的举措。每个个案对应一个单独的客户，并记录各类人口统计和服务用途信息。
- **telco_extra.sav。**该数据文件与 telco.sav 数据文件类似，但删除了 “tenure” 和经对数转换的客户消费变量，代替它们的是标准化的对数转换客户消费变量。
- **telco_missing.sav。** 该数据文件是 telco.sav 数据文件的子集，但某些人口统计数据值已被缺失值替换。
- **testmarket.sav。**该假设数据文件涉及某快餐连锁店为其菜单添加新项目的计划。有三种可能的促销新产品的活动，所以会在多个随机选择的市场中的地点引入新的项目。在每个地点采用不同的促销方式，并记录新项目前四周的每周销售情况。每个个案对应单独地点的一周。

- **testmarket_1month.sav。**该假设数据文件是在数据文件 testmarket.sav 的基础上加上了每周销售“累计”，所以每个个案对应一个单独的地点。所以，一些每周更改的变量消失了，而且现在记录的销售是为期四周的研究过程中的销售之和。
- **tree_car.sav。**该假设数据文件包含人口统计和车辆购买价格数据。
- **tree_credit.sav。**该假设数据文件包含人口统计和银行贷款历史数据。
- **tree_missing_data.sav** 该假设数据文件包含具有大量缺失值的人口统计和银行贷款历史数据。
- **tree_score_car.sav。**该假设数据文件包含人口统计和车辆购买价格数据。
- **tree_textdata.sav。**这是一个只有两个变量的样本数据文件，主要打算在指定测量级别和值标签之前显示变量的默认状态。
- **tv-survey.sav。**该假设数据文件涉及由某电视演播室进行的一项关于是否要继续制作一档成功的节目的调查。906 位调查对象被问及他们在各种情况下是否会收看该节目。每行代表一位单独的调查对象；每列代表一种单独的情况。
- **ulcer_recurrence.sav。**此文件包含某项研究的部分信息，该研究旨在比较两种用来防止溃疡复发的治疗的功效。它提供了区间数据的优秀示例并且已在别处被提出和分析。
- **ulcer_recurrence_recoded.sav。**该文件重新组织 ulcer_recurrence.sav 中的信息以允许为研究的每个区间的事件概率建模而不是简单地为研究结束事件概率建模。它已在别处被提出和分析。
- **verd1985.sav。**该数据文件涉及某项调查。该调查记录了 15 个主体对 8 个变量的响应。需要处理的变量被分成 3 个集。数据集 1 包含 年龄 和 婚姻；数据集 2 包含 宠物 和 新闻；数据集 3 包含 音乐 和 居住。宠物被尺度化为多名义而年龄被尺度化为有序；所有其他变量都被尺度化为单名义。
- **virus.sav。**该假设数据文件涉及某因特网服务提供商 (ISP) 在确定病毒对其网络的影响方面的举措。他们从发现病毒到威胁得以遏制这段时间内跟踪其网络上受感染的电子邮件的流量的（近似）百分比。
- **wheeze_steubenville.sav。**这是关于空气污染对儿童健康影响的纵向研究的一个子集。这些数据包含儿童的气喘状况的重复二分类测量（这些儿童来自 Steubenville，Ohio，年龄为 7 到 10 岁），以及母亲在研究的第一年中是否为吸烟者的固定记录。
- **workprog.sav。**该假设数据文件涉及一份尝试为弱势群体提供较好的工作的政府工作计划。文件后还有一个潜在计划参与者的样本，其中一些参与者是被随机选择来参加该计划的，而其他参与者则不是。每个个案代表一位单独的计划参与者。
- **worldsales.sav** 该假设数据文件包含按不同大洲和产品列出的销售收入。

附录
B

注意事项

这些信息开发用于在全球提供的产品和服务。

IBM 可能在其他国家/地区中不提供在本文档中讨论的产品、服务或功能。请咨询您当地的 IBM 代表以了解有关您所在地区当前可用产品和服务的信息。任何对 IBM 产品、程序或服务的引用，并不意味着仅可使用这些 IBM 产品、程序或服务。作为替代，可以使用任何功能相当的产品、程序或服务，前提是不侵犯任何 IBM 知识产权。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。

在本文档中介绍的主题可能涉及 IBM 的专利或申请中的专利。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.

要查询双字节字符集 (DBCS) 相关许可证信息，请联系所在国家/地区中的 IBM 知识产权部门，或者以书面形式将查询函件发送至：

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

以下段落不适用于英国或此类条款与当地法律不符的其他国家/地区： INTERNATIONAL BUSINESS MACHINES 公司“按原样”提供本出版物，不保证任何明示或暗示，包括但不限于对非侵权性、适销性或对特定用途适用性的暗示担保。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 可能随时对本出版物中所述的产品和/或程序进行改进和/或更改，恕不另行通知。

在本信息材料中对任何非 IBM 网站的引用仅为了方便用户，并不以任何方式表明对这些网站的认可。这些网站上的材料并非本 IBM 产品材料的一部分，您对这些网站的使用需自担风险。

IBM 可以自认为适当并且不会对您构成任何约束的任何方式使用或分发您提供的任何信息。

如果本程序的受许可方试图了解有关程序的信息以启用：(i) 在独立创建的程序和其他程序（包括本程序）之间交换信息；(ii) 相互使用交换的信息，则应联系：

IBM Software Group, Attention:Licensing, 233 S. Wacker Dr., Chicago, IL 60606, USA.

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。

在本文档中介绍的受许可保护程序，及其所有受许可保护材料由 IBM 在双方签署的“IBM 客户协议”、“IBM 国际程序许可证协议”或任何其他等同协议下提供。

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 并未测试过这些产品，无法确认有关非 IBM 产品的性能准确性、兼容性或任何其他声明。有关非 IBM 产品功能的问题应由这些产品的供应商负责。

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。

如果您正在查阅此信息的软拷贝，照片和彩色插图可能不会显示。

商标

IBM、IBM 徽标、ibm.com 和 SPSS 是 IBM Corporation 的商标，在全球许多司法辖区注册。有关最新的 IBM 商标列表，请访问网页 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。

Java 和所有基于 Java 的商标和徽标是 Sun Microsystems, Inc. 在美国、其他国家或地区或两者的商标。

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。

此产品使用 Polar 工程咨询公司的 WinWrap Basic，版权 1993 年-2007 年，http://www.winwrap.com。

其他产品和服务名称可能是 IBM 或其他公司的商标。

Adobe 产品屏幕截图重印已获得 Adobe Systems Incorporated 的许可。

Microsoft 产品屏幕截图重印已获得 Microsoft Corporation 的许可。

索引