

IBM SPSS Direct Marketing 22

IBM

注释

使用本信息及其支持的产品之前，请阅读 第 25 页的『声明』中的信息。

产品信息

此版本适用于 IBM SPSS Statistics V22R0M0 及所有后续发行版和修改版，除非在新版本中另有说明。

目录

第 1 章 直销	1	第 5 章 邮政编码响应率	15
第 2 章 RFM 分析	3	设置	16
来自交易数据的 RFM 得分	3	创建分类响应字段	17
来自客户数据的 RFM 得分	4	第 6 章 购买倾向	19
RFM 分箱化	4	设置	20
保存来自交易数据的 RFM 得分	6	创建分类响应字段	21
保存来自客户数据的 RFM 得分	6	第 7 章 控制包装检验	23
RFM 输出	7	声明	25
第 3 章 聚类分析	9	商标	26
设置	10	索引	29
第 4 章 潜在客户概要文件	11		
设置	12		
创建分类响应字段	12		

第 1 章 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。

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。

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。请参阅 第 11 页的第 4 章，『潜在客户概要文件』主题以获取更多信息。

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。请参阅 第 15 页的第 5 章，『邮政编码响应率』主题以获取更多信息。

购买倾向。 此方法使用试验邮寄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。请参阅 第 19 页的第 6 章，『购买倾向』主题以获取更多信息。

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。请参阅 第 23 页的第 7 章，『控制包装检验』主题以获取更多信息。

第 2 章 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**上次消费时间**。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，为客户分配一个上次消费时间得分。此得分基于将上次消费时间值简单等级排序为少量类别。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上次消费时间等级 5，而拥有过去购买日期的客户将获得上次消费时间等级 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。

结果是每个客户的四个得分：上次消费时间、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。请参阅『来自交易数据的 RFM 得分』主题以获取更多信息。
- 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。请参阅第 4 页的『来自客户数据的 RFM 得分』主题以获取更多信息。

来自交易数据的 RFM 得分

数据注意事项

数据集必须包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。

创建来自交易数据的 RFM 得分

1. 从菜单中选择：

直销 > 选择方法

2. 选择**帮助标识我的最佳联系人 (RFM 分析)**，并单击**继续**。
3. 选择**交易数据**，然后单击**继续**。

4. 选择包含交易日期的变量。
5. 选择包含每次交易的消费金额的变量。
6. 选择汇总每个客户交易金额的方法：总数（所有交易总和）、平均值、中位数或最大值（最高交易金额）。
7. 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。这将用于计算上次消费时间得分。
- 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），但也可能是平均值（平均值）、最大值（最大金额）或其他汇总测量。

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

1. 从菜单中选择：

直销 > 选择方法

2. 选择**帮助标识我的最佳联系人 (RFM 分析)**，并单击**继续**。
3. 选择**客户数据**，然后单击**继续**。
4. 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变量的变量。
5. 选择包含每个客户交易总次数的变量。
6. 选择包含每个客户摘要消费金额的变量。
7. 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 分箱化

将大量数值分组为小量类别的过程有时也称为**分箱化**。在 RFM 分析中，分箱是已评级的类别。您可以使用“分箱化”选项卡修改用于将上次消费时间、频率和消费金额值分配到这些分箱中的方法。

分箱化方法

嵌套。 在嵌套分箱化中，简单等级被分配到上次消费时间值。在每个上次消费时间等级中，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，拥有上次消费时间等级 5 的客户的频率等级 5 与拥有上次消费时间等级 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上次消费时间等级。

独立。 简单等级被分配到上次消费时间、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，无论其上次消费时间得分如何。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。

分箱数

用于每个组件创建 RFM 得分的类别（分箱）数。可能的合并 RFM 得分的总数是三个值的乘积。例如，5 个上次消费时间分箱、4 个频率分箱和 3 个消费金额分箱将创建总共 60 个可能的合并 RFM 得分，范围从 111 到 543。

- 每个组件的缺省分箱数是 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分箱数是 9。

结

“同数”是两个或更多相等的上次消费时间、频率或消费金额值。理想状况下，您希望在每个分箱中拥有大致相同的客户数量，但是大量同数的值可能影响分箱的分布。有两种方法可以处理同数：

- **将同数分配到相同的分箱。** 无论对分箱分布的影响如何，此方法始终将同数的值分配到相同的分箱。此操作提供一致性分箱方法：如果两个客户具有相同的上次消费时间值，那么他们将始终分配到相同的上次消费时间得分。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 分箱评级中，50% 的客户因此获得了上次消费时间得分 5，而非理想值 20%。

注意，使用嵌套分箱化方法时，“一致性”对于频率和消费金额得分有点过于复杂了，因为要在上次消费时间得分分箱中分配频率得分，在频率得分分箱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上次消费时间得分，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分箱的平均分布；因此为了将值分配给已评级的分箱，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。尽管这可以使分箱分布平均（每个分箱中的客户数大致相同），但对于似乎拥有类似或相同上次消费时间、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 1. 将同数分配给相同的分箱以及随机分配同数。

标识	最近购买（上次消费时间）	将同数分配给相同的分箱	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分箱导致分箱分布不均：5 (10%)、4 (40%)、3 (20%)、2 (20%) 和 1 (10%)
- 随机分配同数导致每个分箱中分到 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分箱，并将拥有日期值 8/13/2006 的两个个案也分配给不同的分箱。

注意，将同数分配给不同的分箱的方式完全随机（受结果要和每个分箱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 的上次消费时间等级 5 和个案 5 的上次消费时间等级 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。

变量(A)

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）。
- **上次消费时间得分。** 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汇总测量的得分。得分越高表示消费金额汇总测量的值越高。
- **RFM 得分。** 三个单个得分合为一个值: $(\text{上次消费时间得分} \times 100) + (\text{频率得分} \times 10) + \text{消费金额得分}$ 。

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。

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(A)

选择(选中)想要保存的得分变量:

- **上次消费时间得分。** 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: $(\text{上次消费时间得分} \times 100) + (\text{频率得分} \times 10) + \text{消费金额得分}$ 。

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

已分箱化数据

分箱化数据的图表基于计算的上次消费时间、频率和消费金额得分。

按上次消费时间和频率绘制的消费金额平均值热图。 平均币值分布热图显示由上次消费时间和频率得分定义的类别的平均货币值。颜色越深的区域表示平均货币值越高。

分箱计数图。 分箱计数图表显示选定分箱化方法的分箱分布。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分箱的缺省分箱化方法时，必然会产生一定量的偏差。
- 分箱分布中的极值波动和/或较多空的分箱可能表明您应尝试另一种分箱化方法（分箱数量和/或随机分配结数量较少），或重新考虑 RFM 分析的适用性。

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

未分箱化数据

未分箱化数据的图表基于用来创建上次消费时间、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上次消费时间、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。但对于上次消费时间，图表的解释依赖于上次消费时间测量的类型：日期或时间间隔。对于日期，左侧条代表更“早”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“近”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上次消费时间、频率和货币值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上次消费时间值和数千个货币值。

上次消费时间轴的解释依赖于上次消费时间测量的类型：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“近”的值。

第 3 章 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。

- **名义 (Nominal).** 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 (Ordinal).** 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。

表 2. 测量级别图标

	数值	字符串	日期	时间
刻度（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，因此所有变量都必须都定义有测量级别。

扫描数据。 读取活动数据集中的数据，并分配缺省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

1. 选择**将我的联系人分段到聚类**。
2. 选择要用于创建段的分类（名义、有序）字段和连续（刻度）字段。
3. 单击**运行**以运行该过程。

设置

“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。

第 4 章 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团体。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。 根据试验邮寄的结果，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(O)

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。

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。请参阅 第 12 页的『创建分类响应字段』主题以获取更多信息。

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（刻度）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。

- **名义 (Nominal).** 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 (Ordinal).** 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。

表 3. 测量级别图标

	数值	字符串	日期	时间
刻度（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，因此所有变量都必须定义有测量级别。

扫描数据。 读取活动数据集中的数据，并分配缺省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

1. 选择生成对产品做出响应的我的联系人的概要文件。
2. 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。
3. 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。
4. 选择用于创建概要文件的字段。
5. 单击运行以运行该过程。

设置

“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）。缺省情况下，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（刻度）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:

$$\text{NewName} = \text{OldName} > 0$$

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:

$$\text{NewName}=\text{NOT}(\text{MISSING}(\text{OldName}))$$

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

1. 从菜单中选择:

转换 > 计算变量

2. 为“目标变量”输入新的字段 (变量) 名称。
3. 如果负响应被记录为 0, 则为“数值表达式”输入 $\text{OldName}>0$, 其中 *OldName* 为原始字段名。
4. 如果负响应被记录为缺失 (空), 则为“数值表达式”输入 $\text{NOT}(\text{MISSING}(\text{OldName}))$, 其中 *OldName* 为原始字段名。

第 5 章 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(O)

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。

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 $N \times P \times (1-P)$ 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。
- **等级。** 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。请参阅第 17 页的『创建分类响应字段』主题以获取更多信息。

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价值标签，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

1. 选择标识**最佳响应邮政编码**。
2. 选择标识哪些联系人对产品做出响应的字段。
3. 输入表示正响应的值。如果有任意值定义了值标签，则可以从下拉列表中选择值标签，这将显示对应的值。
4. 选择包含邮政编码的字段。
5. 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

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舍。

输出(O)

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： $(\text{邮寄包装的成本}/\text{每次响应的净收入}) \times 100$ 。这两个值必须为正数。该结果应该是大于 0 并小于 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： $(0.75/56) \times 100 = 1.34\%$ 。

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（刻度）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：

$$\text{NewName}=\text{OldName}>0$$

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：

$$\text{NewName}=\text{NOT}(\text{MISSING}(\text{OldName}))$$

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

1. 从菜单中选择：

转换 > 计算变量

2. 为“目标变量”输入新的字段（变量）名称。
3. 如果负响应被记录为 0，则为“数值表达式”输入 $\text{OldName}>0$ ，其中 *OldName* 为原始字段名。
4. 如果负响应被记录为缺失（空），则为“数值表达式”输入 $\text{NOT}(\text{MISSING}(\text{OldName}))$ ，其中 *OldName* 为原始字段名。

第 6 章 购买倾向

购买倾向使用试验邮寄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。响应字段显示谁对试验邮寄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：

1. 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试验邮寄的结果或来自过去类似活动的响应信息。
2. 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。 公司直销部门使用试验邮寄的结果，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(O)

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总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。

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。请参阅第 21 页的『创建分类响应字段』主题以获取更多信息。

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（刻度）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。

- **名义 (Nominal)**. 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 (Ordinal)**. 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。

表 4. 测量级别图标












	数值	字符串	日期	时间
刻度（连续）		n/a		

表 4. 测量级别图标 (续)

	数值	字符串	日期	时间
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，因此所有变量都必须定义有测量级别。

扫描数据。 读取活动数据集中的数据，并分配缺省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

1. 选择**选择最有可能购买的联系人**。
2. 选择标识哪些联系人对产品做出响应的字段。
3. 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。
4. 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:

5. 选择（勾选）**将模型信息导出到 XML 文件**。
6. 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。
7. 单击**运行**以运行该过程。

使用模型文件对其他数据集评分:

8. 打开您要评分的数据集。
9. 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

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，缺省训练样本大小为 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

总体模型质量。 显示总体模型质量（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缺省名称和标签。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 IBM 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（刻度）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：

$$\text{NewName}=\text{OldName}>0$$

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：

$$\text{NewName}=\text{NOT}(\text{MISSING}(\text{OldName}))$$

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

1. 从菜单中选择:

转换 > 计算变量

2. 为“目标变量”输入新的字段（变量）名称。
3. 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。
4. 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。

第 7 章 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。因此他们发出试验邮寄，以确定新包装能否产生明显更高的正响应率。试验邮寄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(O)

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。

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（刻度）。

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

1. 选择**比较活动效果**。
2. 选择标识每个联系人所属活动组的字段（例如 Offer A、Offer B 等）。该字段必须为名义或有序字段。
3. 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。

4. 单击**运行**以运行该过程。

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 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

[G]

购买倾向 19

[J]

聚类 9

聚类分析 9

聚类分析（直销选项） 9

[K]

控制包装检验 23

[Q]

潜在客户概要文件（直销选项） 11

[Y]

邮政编码响应率 15

L

Logistic 回归（直销） 19

R

RFM 3, 6, 7

分箱化 4

交易数据 3

客户数据 4



Printed in China