

IBM SPSS Direct Marketing 29

IBM

참고

이 정보와 이 정보가 지원하는 제품을 사용하기 전에, [17 페이지의 『주의사항』](#)에 있는 정보를 확인하십시오.

제품 정보

이 개정판은 새 개정판에 별도로 명시하지 않는 한, IBM® SPSS® Statistics 의 버전 29, 릴리스 0, 수정 1 및 모든 후속 릴리스와 수정에 적용됩니다.

© Copyright International Business Machines Corporation .

목차

제 1 장	다이렉트 마케팅	1
다이렉트 마케팅		1
RFM 분석		1
트랜잭션 데이터로부터의 RFM		2
고객 데이터로부터의 RFM 점수		2
RFM 구간화		3
트랜잭션 데이터로부터의 RFM 점수 저장		4
고객 데이터로부터의 RFM 점수 저장		5
RFM 출력		5
군집 분석		6
설정		7
예상 프로파일		7
설정		9
범주형 응답 필드 작성		9
우편번호 응답률		9
설정		10
범주형 응답 필드 작성		11
구매 성향		12
설정		13
범주형 응답 필드 작성		14
패키지 검정 제어		15
주의사항		17
상표		18
색인		19

제 1 장 다이렉트 마케팅

다음 다이렉트 마케팅 기능은 SPSS Statistics Premium Edition 또는 다이렉트 마케팅 옵션에 포함되어 있습니다.

다이렉트 마케팅

직접 마케팅 옵션에는 다양한 사용자 그룹을 정의하고 긍정적인 응답 비율을 최대화하기 위해 특정 그룹을 대상으로 하는 인구 통계, 구매 및 기타 특성을 식별하여 직접 마케팅 캠페인의 결과를 개선하도록 설계된 도구 세트가 포함됩니다.

RFM 분석. 이 방법은 새로운 제안에 응답할 가능성이 가장 높은 기존 고객을 식별합니다.

군집 분석. 군집 분석은 사용하는 데이터 내의 자연적 그룹(또는 군집)을 명확히 나타내도록 설계된 탐색 도구입니다. 예를 들어, 다양한 인구통계적 특성 및 구매 특성을 기반으로 서로 다른 고객 그룹을 식별할 수 있습니다.

예상 프로파일. 이 방법은 이전 캠페인 또는 테스트 캠페인의 결과를 사용하여 설명적 프로파일을 작성합니다. 이 프로파일을 사용하여 향후 캠페인에 특정 연락처 그룹을 대상으로 지정할 수 있습니다. 자세한 정보는 [7 페이지](#)의 『[예상 프로파일](#)』 주제를 참조하십시오.

우편번호 응답률. 이 방법은 이전 캠페인의 결과를 사용하여 우편번호 응답률을 계산합니다. 이러한 비율을 사용하여 향후 캠페인에 특정 우편번호를 대상으로 지정할 수 있습니다. 자세한 정보는 [9 페이지](#)의 『[우편번호 응답률](#)』 주제를 참조하십시오.

구매 성향. 이 방법은 테스트 메일 또는 이전 캠페인의 결과를 사용하여 성향 점수를 생성합니다. 점수는 어느 연락처가 응답할 가능성이 가장 높은지를 나타냅니다. 자세한 정보는 [12 페이지](#)의 『[구매 성향](#)』 주제를 참조하십시오.

제어 패키지 검정. 이 방법은 마케팅 캠페인을 비교하여 서로 다른 패키지 또는 제안의 효과에 유의한 차이가 있는지 확인합니다. 자세한 정보는 [15 페이지](#)의 『[패키지 검정 제어](#)』 주제를 참조하십시오.

RFM 분석

RFM 분석은 새로운 제안에 응답할 가능성이 가장 높은 기존 고객을 식별하는 데 사용되는 방법입니다. 이 기법은 일반적으로 다이렉트 마케팅에 사용됩니다. RFM 분석은 다음과 같은 단순한 이론을 기반으로 합니다.

- 새로운 제안에 응답할 가능성이 가장 높은 고객을 식별하는 데 가장 중요한 요인은 **최근성**입니다. 보다 최근에 구매한 고객은 더 이전에 구매한 고객보다 다시 구매할 가능성이 더 높습니다.
- 두 번째 중요한 요인은 **빈도**입니다. 과거 구매 빈도가 높은 고객이 구매 빈도가 낮은 고객보다 응답할 가능성이 더 높습니다.
- 세 번째 중요한 요인은 소비한 총 금액(**구매총액**이라고 함)입니다. 과거에 보다 많은 금액을 지불한(모든 구매를 합쳐) 고객이 적게 지불한 고객보다 응답할 가능성이 더 높습니다.

RFM 분석 작용 방법

- 가장 최근 구매 날짜 또는 가장 최근 구매 이후의 시간 간격을 기반으로 고객에게 최근성 점수가 할당됩니다. 최근성 값을 단순히 소수의 범주로 분류하고 순위를 지정함으로써 최근성 점수가 결정됩니다. 예를 들어, 5개의 범주를 사용하는 경우, 구매 날짜가 가장 최근인 고객은 최근성 순위가 5가 되고 구매 날짜가 가장 오래된 고객은 최근성 순위가 1이 됩니다.
- 유사한 방식으로 고객에게 빈도 순위가 할당됩니다. 구매 빈도가 높을수록 더 높은 값이 할당됩니다. 예를 들어, 5개의 범주에 순위를 지정하는 경우 가장 빈번하게 구매한 고객의 빈도 순위는 5가 됩니다.
- 끝으로 구매총액 값을 기준으로 고객에게 순위가 할당됩니다. 구매총액 값이 높을수록 순위가 높습니다. 계속해서 5범주를 예로 들 경우, 가장 많이 지불한 고객은 구매총액 순위가 5가 됩니다.

결과적으로 고객마다 4개의 점수를 가집니다. 최근성, 빈도, 구매총액 및 결합 RFM 점수이며, 결합 RFM 점수는 단순히 세 개의 개별 점수를 하나의 값으로 연결한 것입니다. "최고"의 고객(제안에 응답할 가능성이 가장 높은

고객)은 결합 RFM 점수가 가장 높은 고객입니다. 예를 들어, 5범주 순위 지정에서는 총 125가지의 결합 RFM 점수가 가능하고 가장 높은 결합 RFM 점수는 555입니다.

데이터 고려사항

- 데이터 행이 트랜잭션을 나타내는 경우(각 행이 단일 트랜잭션을 나타내고 고객마다 여러 트랜잭션이 있을 수 있음) 트랜잭션으로부터의 RFM을 사용합니다. 자세한 정보는 2 페이지의 『트랜잭션 데이터로부터의 RFM』 주제를 참조하십시오.
- 데이터 행이 고객과 함께 모든 트랜잭션에 대한 요약 정보를 나타내는 경우(소비한 총액, 총 트랜잭션 수 및 가장 최근 트랜잭션 날짜에 대한 값을 나타내는 열 포함) 고객 데이터로부터의 RFM을 사용합니다. 자세한 정보는 2 페이지의 『고객 데이터로부터의 RFM 점수』 주제를 참조하십시오.

트랜잭션 데이터로부터의 RFM

데이터 고려사항

데이터 세트에는 다음 정보가 포함된 변수가 있어야 합니다.

- 각 케이스(고객)를 식별하는 변수 또는 변수 조합
- 각 트랜잭션의 날짜가 포함된 변수
- 각 트랜잭션의 구매총액 값이 포함된 변수

트랜잭션 데이터에서 RFM 점수 작성

1. 메뉴에서 다음을 선택합니다.

다이렉트 마케팅 > 기술 선택

2. **최고의 연락처 식별 지원(RFM 분석)**을 선택하고 **계속**을 클릭합니다.
3. **트랜잭션 데이터**를 선택하고 **계속**을 클릭합니다.
4. 트랜잭션 날짜가 포함된 변수를 선택합니다.
5. 각 트랜잭션의 구매 금액이 포함된 변수를 선택합니다.
6. 각 고객의 트랜잭션 양을 요약하는 방법을 총계(모든 트랜잭션의 합), 평균값, 중앙값 또는 최대값(가장 높은 트랜잭션 양)에서 선택합니다.
7. 각 고객을 고유하게 식별하는 변수 또는 변수의 조합을 선택합니다. 예를 들어, 고유 ID 코드를 사용하거나 성과 이름을 조합하여 케이스를 식별할 수 있습니다.

고객 데이터로부터의 RFM 점수

데이터 고려사항

데이터 세트에는 다음 정보가 포함된 변수가 있어야 합니다.

- 가장 최근 구매 날짜 또는 가장 최근 구매 날짜 이후의 시간 간격. 이 정보는 최근성 점수를 계산하는 데 사용됩니다.
- 총 구매 수. 이 정보는 빈도 점수를 계산하는 데 사용됩니다.
- 모든 구매의 구매총액 집계값. 이 정보는 구매총액 점수를 계산하는 데 사용됩니다. 일반적으로 이 값은 모든 구매를 합한 값(총계)이지만 평균, 최대값 또는 기타 집계값일 수도 있습니다.

새 데이터 세트에 RFM 점수를 쓰려는 경우, 활성 데이터 세트에도 각 케이스(고객)를 식별하는 변수 또는 변수 조합이 있어야 합니다.

고객 데이터에서 RFM 점수 작성

1. 메뉴에서 다음을 선택합니다.

다이렉트 마케팅 > 기술 선택

2. **최고의 연락처 식별 지원(RFM 분석)**을 선택하고 **계속**을 클릭합니다.
3. **고객 데이터**를 선택하고 **계속**을 클릭합니다.

4. 가장 최근 트랜잭션 날짜 또는 가장 최근 트랜잭션 이후 시간 간격을 나타내는 숫자가 포함된 변수를 선택합니다.
5. 각 고객의 총 트랜잭션 수가 포함된 변수를 선택합니다.
6. 각 고객의 구매총액 집계값이 포함된 변수를 선택합니다.
7. 새 데이터 세트에 RFM 점수를 쓰려는 경우, 각 고객을 고유하게 식별하는 변수 또는 변수 조합을 선택합니다. 예를 들어, 고유 ID 코드를 사용하거나 성과 이름을 조합하여 케이스를 식별할 수 있습니다.

RFM 구간화

다수의 숫자 값을 소수의 범주로 그룹화하는 프로세스를 **구간화**라고도 합니다. RFM 분석에서 구간은 순위가 지정된 범주입니다. 구간화 탭을 사용하여 최근성, 빈도 및 구매총액 값을 이러한 구간에 할당하는 데 사용되는 방법을 수정할 수 있습니다.

구간화 방법

중첩. 중첩 구간화에서는 최근성 값에 단순한 순위가 할당됩니다. 각 최근성 순위 내에서 고객에게 빈도 순위가 할당되고 각 빈도 순위 내에서 고객에게 구매총액 순위가 할당됩니다. 이 방법을 사용하는 경우 결합 RFM 점수가 보다 균등하게 분포되는 경향이 있지만 빈도 및 구매총액 순위 점수를 해석하기가 더욱 어려워지는 단점이 있습니다. 예를 들어, 빈도 순위는 최근성 순위에 의존하므로 최근성 순위가 5인 고객의 빈도 순위 5와 최근성 순위가 4인 고객의 빈도 순위 5는 동일한 것을 의미하지 않을 수도 있습니다.

독립. 최근성, 빈도 및 구매총액 값에 단순한 순위가 할당됩니다. 세 개의 순위가 독립적으로 할당됩니다. 따라서 세 RFM 구성요소 각각의 해석이 명확합니다. 즉, 고객의 최근성 점수와 관계 없이 한 고객의 빈도 점수 5는 다른 고객의 빈도 점수 5와 동일함을 의미합니다. 보다 작은 표본에서는 결합 RFM 점수가 균등하지 않게 분포되는 단점이 있습니다.

구간 수

각 구성요소에서 RFM 점수를 작성하기 위해 사용할 범주(구간)의 수입니다. 가능한 결합 RFM 점수의 총 수는 세 값을 곱한 값입니다. 예를 들어, 5개의 최근성 구간, 4개의 빈도 구간 및 3개의 구매총액 구간은 111에서 543까지 총 60개의 결합 RFM 점수를 작성합니다.

- 각 구성요소의 기본값은 5이며 이 경우에 작성되는 결합 RFM 점수는 111에서 555까지 125개입니다.
- 각 점수 구성요소에 허용되는 구간의 최대 수는 9입니다.

등순위

"등순위"는 단순히 둘 이상의 동일한 최근성, 빈도 또는 구매총액 값입니다. 각 구간에 거의 동일한 수의 고객이 포함되는 것이 이상적이지만 등순위 값이 다수인 경우 구간 분포에 영향을 미칠 수 있습니다. 등순위를 처리하는 두 가지 방법이 있습니다.

- **등순위를 동일한 구간에 할당.** 이 방법은 구간 분포에 미치는 영향과 관계 없이 항상 동일한 구간에 등순위 값을 할당합니다. 이 방법은 일관된 구간화 방법을 제공합니다. 두 고객의 최근성 값이 동일한 경우 두 고객에게는 항상 동일한 최근성 점수가 할당됩니다. 그러나 극단적인 예로 1000명의 고객 중 500명의 가장 최근 구매 날짜가 동일할 수 있습니다. 이 경우 5구간 순위 지정에서는 이상적인 값인 20% 대신 50%의 고객이 최근성 점수 5를 취득합니다.

중첩 구간화 방법에서는 빈도 점수가 최근성 점수 구간 내에서 할당되고 구매총액 점수가 빈도 점수 구간 내에서 할당되므로 빈도 및 구매총액 점수의 경우 "일관성"이 좀 더 복잡해집니다. 따라서 빈도 값이 동일한 두 고객의 최근성 점수가 다른 경우에는 등순위 값이 처리되는 방법과 무관하게 두 고객의 빈도 점수가 다를 수 있습니다.

- **등순위를 무작위로 할당.** 이 방법을 사용하는 경우, 순위를 지정하기 전에 등순위에 매우 작은 무작위 분산 요인을 할당함으로써 구간 분포가 균등해집니다. 즉, 순위 지정된 구간에 값을 할당하기 위해 등순위 값이 없어집니다. 이 프로세스는 원래의 값에는 아무런 영향을 미치지 않습니다. 등순위를 명확히 구별하는 데에만 사용됩니다. 이 방법을 사용하는 경우 구간 분포가 균등(각 구간에 거의 동일한 수의 고객이 포함됨)해지는 반면, 최근성, 빈도 및/또는 구매총액 값이 유사하거나 동일해 보이는 고객들의 점수 결과가 완전히 다를 수 있습니다(특히, 총 고객 수가 비교적 작거나 등순위 수가 비교적 높은 경우).

ID	가장 최근의 구매(최근성)	동순위를 동일한 구간에 할당	동순위를 무작위로 할당
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

- 이 예에서 동일한 구간에 동순위를 할당하는 경우 구간 분포가 불균등해집니다(5(10%), 4(40%), 3(20%), 2(20%), 1(10%)).
- 동순위를 무작위로 할당하는 경우 각 구간에 20%씩 분포되지만, 이러한 결과를 얻기 위해 날짜 값이 10/28/2006인 4개의 케이스가 서로 다른 3개의 구간에 할당되고 날짜 값이 8/13/2006인 2개의 케이스가 서로 다른 구간에 할당됩니다.

동순위가 서로 다른 구간에 할당되는 방법은 완전히 무작위입니다(최종 결과는 각 구간에 동일한 수의 케이스가 포함된다는 제한조건 내에서). 동일한 방법을 사용하여 다른 점수 세트를 계산하는 경우 동순위 값을 갖는 특정 케이스에 대한 순위 지정이 달라질 수 있습니다. 예를 들어, 케이스 4의 최근성 순위가 5이고 케이스 5의 최근성 순위가 3인 경우 다음 번에 그 순위가 서로 바뀔 수 있습니다.

트랜잭션 데이터로부터의 RFM 점수 저장

트랜잭션 데이터로부터의 RFM은 항상 고객당 하나의 행이 있는 집계된 새 데이터 세트를 작성합니다. 저장 탭을 사용하여 저장할 점수 및 기타 변수와 이들의 저장 위치를 지정하십시오.

변수

각 고객을 고유하게 식별하는 ID 변수는 새 데이터 세트에 자동으로 저장됩니다. 새 데이터 세트에 다음과 같은 추가 변수를 저장할 수 있습니다.

- **각 고객의 가장 최근 트랜잭션 날짜.**
- **트랜잭션 수.** 각 고객의 트랜잭션 행 총 수.
- **금액.** 변수 탭에서 선택한 집계 방법에 기반한 각 고객의 집계 금액입니다.
- **최근성 점수.** 가장 최근 트랜잭션 날짜를 기준으로 각 고객에게 할당되는 점수입니다. 점수가 높을수록 트랜잭션 날짜가 더 최근입니다.
- **빈도 점수.** 총 트랜잭션 수를 기준으로 각 고객에게 할당되는 점수입니다. 점수가 높을수록 트랜잭션 수가 더 많습니다.
- **구매총액 점수.** 선택한 구매총액 집계값을 기준으로 각 고객에게 할당되는 점수입니다. 점수가 높을수록 구매총액 집계값이 높습니다.
- **RFM 점수.** 세 개의 개별 점수를 하나의 값으로 결합한 것입니다. $(\text{최근성} \times 100) + (\text{빈도} \times 10) + \text{구매총액}$.

기본적으로 사용 가능한 모든 변수가 새 데이터 세트에 포함됩니다. 따라서 제외시킬 변수를 선택 취소하십시오. 또는 고유의 변수 이름을 지정할 수 있습니다. 변수 이름은 표준 변수 이름 지정 규칙을 따라야 합니다.

위치

트랜잭션 데이터로부터의 RFM은 항상 고객당 하나의 행이 있는 집계된 새 데이터 세트를 작성합니다. 현재 세션에서 새 데이터 세트를 작성하거나 RFM 점수 데이터를 외부 데이터 파일에 저장할 수 있습니다. 데이터 세트 이름은 표준 변수 이름 지정 규칙을 따라야 합니다. (이 제한사항은 외부 데이터 파일 이름에는 적용되지 않습니다.)

고객 데이터로부터의 RFM 점수 저장

고객 데이터의 경우, RFM 점수 변수를 활성 데이터 세트에 추가하거나 선택한 점수 변수가 포함된 새 데이터 세트를 작성할 수 있습니다. 저장 탭을 사용하여 저장할 점수 변수와 이들의 저장 위치를 지정하십시오.

저장된 변수의 이름

- **고유 이름 자동 생성.** 이 옵션을 사용하면 활성 데이터 세트에 점수 변수를 추가할 때 새 변수 이름이 고유합니다. 이 옵션은 서로 다른 여러 RFM 점수 세트(서로 다른 기준에 기반한)를 활성 데이터 세트에 추가하는 경우에 특히 유용합니다.
- **사용자 정의 이름.** 점수 변수에 고유의 변수 이름을 할당하는 데 사용됩니다. 변수 이름은 표준 변수 이름 지정 규칙을 따라야 합니다. 주제를 참조하십시오.

변수

저장할 점수 변수를 선택하십시오.

- **최근성 점수.** 변수 탭에서 선택한 트랜잭션 날짜 또는 간격 변수의 값을 기준으로 각 고객에게 할당되는 점수입니다. 날짜가 더 최근일수록 또는 간격 값이 낮을수록 더 높은 점수가 할당됩니다.
- **빈도 점수.** 변수 탭에서 선택한 트랜잭션 수 변수를 기준으로 각 고객에게 할당되는 점수입니다. 값이 높을수록 더 높은 점수가 할당됩니다.
- **구매총액 점수.** 변수 탭에서 선택한 금액 변수를 기준으로 각 고객에게 할당되는 점수입니다. 값이 높을수록 더 높은 점수가 할당됩니다.
- **RFM 점수.** 세 개의 개별 점수를 하나의 값으로 결합한 것입니다. $(\text{최근성} * 100) + (\text{빈도} * 10) + \text{구매총액}$.

위치

고객 데이터의 경우, 다음과 같은 세 가지 위치에 새 RFM 점수를 저장할 수 있습니다.

- **활성 데이터 세트.** 선택한 RFM 점수 변수가 활성 데이터 세트에 추가됩니다.
- **새 데이터 세트.** 선택한 RFM 점수 변수와 각 고객(케이스)을 고유하게 식별하는 ID 변수가 현재 세션의 새 데이터 세트에 기록됩니다. 데이터 세트 이름은 표준 변수 이름 지정 규칙을 따라야 합니다. 주제를 참조하십시오. 이 옵션은 변수 탭에서 하나 이상의 고객 식별자 변수를 선택하는 경우에만 사용할 수 있습니다.
- **파일.** 선택한 RFM 점수와 각 고객(케이스)을 고유하게 식별하는 ID 변수가 외부 데이터 파일에 저장됩니다. 이 옵션은 변수 탭에서 하나 이상의 고객 식별자 변수를 선택하는 경우에만 사용할 수 있습니다.

RFM 출력

구간화된 데이터

구간화된 데이터의 도표 및 표는 계산된 최근성, 빈도 및 구매총액 점수를 기반으로 합니다.

최근성 및 빈도별 평균 구매총액 값 히트 맵. 평균 통화 분포의 히트 맵은 최근성 점수 및 구매빈도 점수별로 정의된 범주의 평균 통화 값을 보여줍니다. 어두운 영역일수록 평균 통화 값이 높습니다.

구간 개수 도표. 빈도 도표는 선택한 구간화 방법의 구간 분포를 표시합니다. 각 막대는 각 결합 RFM 점수가 할당될 케이스의 수를 나타냅니다.

- 거의 같은 높이의 모든(또는 대부분) 막대가 보통 고르게 분포되기를 바라지만, 동률 값을 같은 구간에 할당하는 기본 구간화 방법을 사용할 경우 분산의 특정 양이 예상됩니다.
- 구간 분포의 극단적인 변동이 있거나 빈 구간이 많은 경우 다른 구간화 방법(더 적은 수의 구간 사용 및/또는 동순위의 무작위 할당)을 시도하거나 RFM 분석이 적합한지 다시 검토해야 합니다.

구간 개수 테이블. 구간 개수 도표에 있는 정보와 동일한 정보를 테이블 양식으로 표시하며 각 셀에 구간 개수가 포함됩니다.

구간화되지 않은 데이터

구간화되지 않은 데이터의 도표 및 테이블은 최근성, 빈도 및 구매총액 점수를 작성하는 데 사용되는 원래 변수를 기반으로 합니다.

히스토그램. 히스토그램은 최근성, 빈도 및 구매총액 점수를 계산하는 데 사용되는 세 가지 변수 값의 상대분포를 보여줍니다. 이러한 히스토그램은 흔히 정규분포나 대칭분포보다는 다소 기운 분포를 나타냅니다.

각 히스토그램의 수평축은 항상 낮은 값이 왼쪽에 있고 오른쪽으로 갈수록 값이 높아집니다. 하지만 최근의 경우 도표 해석은 날짜 또는 시간 간격과 같은 최근성 측도의 유형에 따라 다릅니다. 날짜의 경우 왼쪽 막대가 훨씬 과거의 값을 나타냅니다(과거에 가까운 날짜가 최근 날짜보다 값이 낮습니다). 시간 구간의 경우 왼쪽 막대가 최근 값을 나타냅니다(시간 구간이 작을수록 최근 트랜잭션입니다).

변수 쌍의 산점도. 이러한 산점도는 최근성, 빈도 및 구매총액 점수를 계산하는 데 사용된 세 가지 변수 사이의 관계를 보여줍니다.

빈도는 비교적 작은 범위의 이산값을 나타내는 경우가 많으므로 빈도 척도에서 점들이 명확한 선형으로 그룹화 되는 것이 일반적입니다. 예를 들어, 총 트랜잭션 수가 15를 초과하지 않는 경우 가능한 빈도 값은 15개뿐이지만(분수 트랜잭션을 포함하지 않는 한) 최근성 값은 수백 개가 될 수 있고 구매총액 값은 수천 개가 될 수 있습니다.

최근성 축의 해석은 날짜 또는 시간 간격과 같은 최근성 측도의 유형에 따라 다릅니다. 날짜의 경우 원점에 가까울수록 과거의 더 오래된 날짜를 나타냅니다. 시간 간격의 경우 원점에 가까울수록 더 최근의 값을 나타냅니다.

군집 분석

군집 분석은 해당 데이터 내에서 자연 그룹(또는 군집)을 나타내기 위해 설계된 탐색 도구입니다. 예를 들어, 다양한 인구통계적 특성 및 구매 특성을 기반으로 서로 다른 고객 그룹을 식별할 수 있습니다.

예. 소매 및 소비자 회사는 고객의 구매 습관, 성별, 나이, 소득 수준 등을 설명하는 데이터에 클러스터링 기법을 정기적으로 적용합니다. 이 회사는 판매를 증가시키고 브랜드 로열티를 구축하기 위해 해당 마케팅 및 제품 개발 전략을 각 고객 그룹에 맞게 조정합니다.

군집 분석 데이터 고려사항

데이터. 이 프로시저는 연속형 변수와 범주형 변수에 모두 적용됩니다. 각 레코드(행)는 군집화할 고객을 나타내고 필드(변수)는 군집화 기준이 되는 속성을 나타냅니다.

레코드 순서. 레코드 순서에 따라 결과가 달라질 수 있습니다. 순서가 미치는 영향을 최소화하려면 레코드의 순서를 무작위로 지정할 수도 있습니다. 주어진 해법의 안정성을 확인하기 위해 레코드를 서로 다른 무작위 순서로 정렬하여 분석을 여러 번 실행할 수도 있습니다.

측정 수준. 측정 수준은 결과의 계산에 영향을 미치므로 올바른 측정 수준을 할당하는 것이 중요합니다.

- **명목(Nominal).** 변수의 값이 고유한 순위가 없는 범주(예: 직원이 근무하는 회사의 부서)를 나타내면 변수를 명목으로 취급할 수 있습니다. 종교, 우편번호 또는 종교 단체 등이 명목변수에 해당합니다.
- **순서(Ordinal).** 변수의 값이 일부 고유한 순위(예: 높은 만족도에 대한 높은 만족도에서 서비스 만족도의 레벨)를 나타내는 범주를 나타내는 경우 변수를 서수로 취급할 수 있습니다. 순서변수의 예로는 만족도나 신뢰도를 나타내는 태도 스코어 및 선호도 등급 스코어가 있습니다.
- **연속.** 변수는 값이 의미 있는 메트릭을 사용하여 정렬된 범주를 나타내므로 값 간 거리 비교가 적절할 때 척도(연속형)로 처리될 수 있습니다. 척도변수의 예로는 연령과 수입이 있습니다.

각 필드 옆의 아이콘은 현재 측정 수준을 나타냅니다.

	숫자	문자열	날짜	시간
척도(연속형)		해당사항 없음		
순서				
명목				

You can change the measurement level in Variable View of the Data Editor or you can use the Define Variable Properties dialog to suggest an appropriate measurement level for each field.

측정 수준을 알 수 없는 필드

측정 수준 경고는 데이터 세트에서 하나 이상의 변수(필드)에 대해 측정 수준을 알 수 없을 때 표시됩니다. 측정 수준은 이 프로시저의 계산 결과에 영향을 미치기 때문에 모든 변수에 대해 측정 수준을 정의해야 합니다.

데이터 스캔. 활성 데이터 세트의 데이터를 읽고 현재 알 수 없는 측정 수준이 있는 필드에 기본 측정 수준을 할당합니다. 데이터 세트가 큰 경우 시간이 걸릴 수 있습니다.

수동으로 할당. 알 수 없는 측정 수준이 있는 필드를 모두 나열하는 대화 상자를 엽니다. 이 대화 상자에서 해당 필드에 측정 수준을 할당할 수 있습니다. 데이터 편집기의 변수 보기에서도 측정 수준을 할당할 수 있습니다.

이 프로시저에 대해 측정 수준이 중요하기 때문에 모든 필드에 정의된 측정 수준이 있을 때까지는 대화 상자에 액세스하여 이 프로시저를 실행할 수 없습니다.

군집 분석 구하기

메뉴에서 다음을 선택합니다.

다이렉트 마케팅 > 기술 선택

1. **연락처를 군집으로 세그먼트화**를 선택합니다.
2. 세그먼트 작성에 사용할 범주형(명목형, 순서형) 및 연속형(척도) 필드를 선택합니다.
3. 프로시저를 실행하려면 **실행**을 클릭하십시오.

설정

설정 탭에서는 세그먼트를 설명하는 도표 및 표를 표시하거나 억제하고 데이터 세트에 있는 각 레코드의 세그먼트(군집)를 식별하는 새 필드를 데이터 세트에 저장하며 군집 해법에 포함시킬 세그먼트 수를 지정할 수 있습니다.

도표 및 테이블 표시. 세그먼트를 설명하는 표 및 도표를 표시합니다.

세그먼트 멤버십. 각 레코드가 속하는 세그먼트를 식별하는 새 필드(변수)를 저장합니다.

- 필드 이름은 IBM SPSS Statistics 이름 지정 규칙을 따라야 합니다. 주제를 참조하십시오.
- 세그먼트 멤버십 필드 이름은 데이터 세트에 이미 있는 필드 이름과 중복되지 않아야 합니다. 동일한 데이터 세트에 이 프로시저를 두 번 이상 실행하는 경우 매번 다른 이름을 지정해야 합니다.
- **세그먼트 수.** 세그먼트 수가 결정되는 방법을 제어합니다.
- **자동 결정.** 프로시저가 "최적"의 세그먼트 수를 지정된 최대값까지 자동으로 결정합니다.

고정 수 지정. 프로시저가 지정된 수의 세그먼트를 생성합니다.

예상 프로파일

이 방법은 이전 캠페인 또는 테스트 캠페인의 결과를 사용하여 설명적 프로파일을 작성합니다. 이 프로파일을 사용하여 향후 캠페인에 특정 연락처 그룹을 대상으로 지정할 수 있습니다. 응답 필드는 이전 또는 테스트 캠페인에 응답한 사람을 나타냅니다. 프로파일 목록에는 프로파일을 작성하는 데 사용할 특성이 있습니다.

예. 한 회사의 다이렉트 마케팅 부서가 테스트 메일의 결과를 기반으로 인구통계적 정보를 사용하여 제안에 응답할 가능성이 가장 높은 고객 유형의 프로파일을 생성하려고 합니다.

결과

출력에는 각 프로파일 그룹에 대한 설명을 제공하고 응답률(긍정 응답의 백분율) 및 누적 응답률을 표시하는 테이블과 누적 응답률의 도표가 포함됩니다. 대상 최소 응답률을 포함시킬 경우, 테이블에는 최소 누적 응답률을 충족하는 프로파일이 색상으로 구별되어 표시되고 도표에는 지정된 최소 응답률 값에서 참조선이 표시됩니다.

예상 프로파일 데이터 고려사항

응답 필드. 응답 필드는 명목형 또는 순서형이어야 합니다. 문자열 또는 숫자일 수 있습니다. 이 필드에 구매 수도 또는 구매 금액을 표시하는 값이 포함되는 경우에는 단일 값이 모든 긍정 응답을 나타내는 새 필드를 작성해야 합니다. 자세한 정보는 9 페이지의 『범주형 응답 필드 작성』 주제를 참조하십시오.












긍정 응답 값. 긍정 응답 값은 긍정으로 응답한(예를 들어, 구매) 고객을 식별합니다. 기타 모든 비결측 응답 값은 부정 응답을 나타내는 것으로 간주됩니다. 응답 필드에 대해 정의된 값 레이블이 있는 경우 그러한 레이블은 드롭 다운 목록에 표시됩니다.

프로파일 작성 도구. 이러한 필드는 명목형, 순서형 또는 연속형(척도)일 수 있습니다. 문자열 또는 숫자일 수 있습니다.

측정 수준. 측정 수준은 결과의 계산에 영향을 미치므로 올바른 측정 수준을 할당하는 것이 중요합니다.

- **명목(Nominal).** 변수의 값이 고유한 순위가 없는 범주(예: 직원이 근무하는 회사의 부서)를 나타내면 변수를 명목으로 취급할 수 있습니다. 종교, 우편번호 또는 종교 단체 등이 명목변수에 해당합니다.
- **순서(Ordinal).** 변수의 값이 일부 고유한 순위(예: 높은 만족도에 대한 높은 만족도에서 서비스 만족도의 레벨)를 나타내는 범주를 나타내는 경우 변수를 서수로 취급할 수 있습니다. 순서변수의 예로는 만족도나 신뢰도를 나타내는 태도 스코어 및 선호도 등급 스코어가 있습니다.
- **연속.** 변수는 값이 의미 있는 매트릭을 사용하여 정렬된 범주를 나타내므로 값 간 거리 비교가 적절할 때 척도(연속형)로 처리될 수 있습니다. 척도변수의 예로는 연령과 수입이 있습니다.

각 필드 옆의 아이콘은 현재 측정 수준을 나타냅니다.

	숫자	문자열	날짜	시간
척도(연속형)		해당사항 없음		
순서				
명목				

You can change the measurement level in Variable View of the Data Editor or you can use the Define Variable Properties dialog to suggest an appropriate measurement level for each field.

측정 수준을 알 수 없는 필드

측정 수준 경고는 데이터 세트에서 하나 이상의 변수(필드)에 대해 측정 수준을 알 수 없을 때 표시됩니다. 측정 수준은 이 프로시저의 계산 결과에 영향을 미치기 때문에 모든 변수에 대해 측정 수준을 정의해야 합니다.

데이터 스캔. 활성 데이터 세트의 데이터를 읽고 현재 알 수 없는 측정 수준이 있는 필드에 기본 측정 수준을 할당합니다. 데이터 세트가 큰 경우 시간이 걸릴 수 있습니다.

수동으로 할당. 알 수 없는 측정 수준이 있는 필드를 모두 나열하는 대화 상자를 엽니다. 이 대화 상자에서 해당 필드에 측정 수준을 할당할 수 있습니다. 데이터 편집기의 변수 보기에서도 측정 수준을 할당할 수 있습니다.

이 프로시저에 대해 측정 수준이 중요하기 때문에 모든 필드에 정의된 측정 수준이 있을 때까지는 대화 상자에 액세스하여 이 프로시저를 실행할 수 없습니다.

예상 프로파일 구하기

메뉴에서 다음을 선택합니다.

다이렉트 마케팅 > 기술 선택

1. 제안에 응답한 연락처의 프로파일 생성을 선택합니다.
2. 제안에 응답한 연락처를 식별하는 필드를 선택합니다. 이 필드는 명목형 또는 순서형이어야 합니다.
3. 긍정 응답을 나타내는 값을 입력합니다. 값에 정의된 값 레이블이 있는 경우, 드롭 다운 목록에서 값 레이블을 선택하면 해당 값이 표시됩니다.
4. 프로파일 작성에 사용할 필드를 선택합니다.
5. 실행을 클릭하여 프로시저를 실행합니다.

설정

설정 탭에서는 최소 프로파일 그룹 크기를 제어하고 출력에 최소 응답률 임계값을 포함시킬 수 있습니다.

최소 프로파일 그룹 크기. 각 프로파일은 데이터 세트에 있는 연락처 그룹의 공유 특성을 나타냅니다(예: 서부에서 사는 40세 미만의 여성). 기본적으로 가장 작은 프로파일 그룹 크기는 100입니다. 그룹 크기가 작을수록 더 많은 그룹을 나타낼 수 있지만 그룹 크기가 클수록 결과의 신뢰성이 더 높습니다. 값은 양의 정수여야 합니다.

결과에 최소 응답률 임계값 정보 포함. 결과에 응답률(긍정 응답의 백분율) 및 누적 응답률을 표시하는 테이블과 누적 응답률의 도표가 포함됩니다. 대상 최소 응답률을 입력하는 경우, 표에는 최소 누적 응답률을 충족하는 프로파일이 색상으로 구별되어 표시되고 도표에는 지정된 최소 응답률 값에서 참조선이 표시됩니다. 값은 0보다 크고 100보다 작아야 합니다.

범주형 응답 필드 작성

응답 필드는 하나의 값이 모든 긍정 응답을 나타내는 범주형 필드여야 합니다. 기타 비결측값은 부정 응답으로 간주됩니다. 응답 필드가 연속형(척도) 값(예: 구매 수 또는 구매 금액)을 나타내는 경우, 0이 아닌 모든 응답값에 단일의 긍정 응답값을 할당하는 새 필드를 작성해야 합니다.

- 부정 응답이 0(공백이 아님. 공백은 결측으로 처리됨)으로 기록되는 경우 이는 다음 공식으로 계산할 수 있습니다.

$newName = OldName > 0$

여기서 *NewName*은 새 필드의 이름이고 *OldName*은 원래 필드의 이름입니다. 이것은 0보다 큰 모든 비결측값에 1 값을 할당하고 0 이하의 모든 비결측값에 0을 할당하는 논리식입니다.

- 부정 응답에 대해 아무 값도 기록되지 않는 경우 이러한 값은 결측값으로 처리되며 공식은 좀 더 복잡합니다.

$newName = NOT(MISSING(OldName))$

이 논리식에서는 모든 비결측 응답값에 1 값이 할당되고 모든 결측 응답값에 0 값이 할당됩니다.

- 부정(0) 응답값과 결측값을 구별할 수 없는 경우에는 정확한 응답값을 계산할 수 없습니다. 실제 결측값 수가 비교적 적은 경우에는 계산되는 응답률에 큰 영향을 미치지 않을 수도 있습니다. 그러나 결측값 수가 많은 경우, 예를 들어, 전체 데이터 세트 중 작은 검정용 표본에 대해서만 응답 정보가 기록되는 경우, 계산된 응답률은 실제 응답률보다 현저히 낮기 때문에 의미가 없게 됩니다.

범주형 응답 필드 작성

1. 메뉴에서 다음을 선택합니다.

변환 > 변수 계산

2. 대상 변수에 새 필드(변수) 이름을 입력합니다.
3. 부정 응답이 0으로 기록되는 경우, 수식에 $OldName > 0$ 을 입력합니다. 여기서 *OldName*은 원래 필드 이름입니다.
4. 부정 응답이 결측값(공백)으로 기록되는 경우 수식에 $NOT(MISSING(OldName))$ 을 입력합니다. 여기서 *OldName*은 원래 필드 이름입니다.

우편번호 응답률

이 방법은 이전 캠페인의 결과를 사용하여 우편번호 응답률을 계산합니다. 이러한 비율을 사용하여 향후 캠페인에 특정 우편번호를 대상으로 지정할 수 있습니다. 반응 필드는 이전 캠페인에서 누가 반응하였는가를 나타냅니다. 지역코드 필드는 지역코드를 포함하고 있는 필드여야 합니다.

예. 한 회사의 다이렉트 마케팅 부서가 이전 메일의 결과를 기반으로 우편번호별로 응답률을 생성하려고 합니다. 사용 가능한 최소 응답률 및/또는 메일에 포함시킬 연락처 최대 수 등의 다양한 기준을 기반으로 특정 우편번호를 대상으로 지정할 수 있습니다.

출력

이 프로시저의 출력에는 우편번호별로 응답 비율을 포함하는 새 데이터 세트 및 십분위 순위(최상위 10%, 최상위 20% 등)별로 결과를 요약하는 테이블과 차트가 포함됩니다. 테이블은 사용자 지정 최소 누적 응답률 또는 최대 연락처 수를 기준으로 색상을 구분할 수 있습니다.

새 데이터 세트에는 다음 필드가 포함됩니다.

- **우편번호.** 우편번호 그룹이 전체 값 중 일부만을 기준으로 하는 경우 이 값은 우편번호의 해당 부분의 값입니다. Excel 파일에서 이 열의 헤더 행 레이블은 원래 데이터 세트에 있는 우편번호 필드의 이름입니다.
- **응답률.** 각 우편번호에서 긍정 응답의 백분율입니다.
- **반응.** 각 우편번호에서 긍정 응답의 수입니다.
- **연락처.** 각 우편번호에서 응답 필드에 비결측값이 포함된 총 연락처 수입니다.
- **색인.** $N \times P \times (1-P)$ 수식에 기반한 "가중치가 부여된" 응답입니다. 여기서 N 은 연락처 수이고 P 는 비율로 표현된 응답률입니다.
- **순위.** 누적 우편번호 응답률의 십분위 순위(최상위 10%, 최상위 20% 등)가 내림차순으로 표시됩니다.

우편번호 응답률 데이터 고려사항

응답 필드. 응답 필드는 문자열 또는 숫자일 수 있습니다. 이 필드에 구매 수 또는 구매 금액을 표시하는 값이 포함되는 경우에는 단일 값이 모든 긍정 응답을 나타내는 새 필드를 작성해야 합니다. 자세한 정보는 [11 페이지의 『범주형 응답 필드 작성』](#) 주제를 참조하십시오.

긍정 응답 값. 긍정 응답 값은 긍정으로 응답한(예를 들어, 구매) 고객을 식별합니다. 기타 모든 비결측 응답 값은 부정 응답을 나타내는 것으로 간주됩니다. 응답 필드에 대해 정의된 값 레이블이 있는 경우 그러한 레이블은 드롭 다운 목록에 표시됩니다. 주제를 참조하십시오.

우편번호 필드. 우편번호 필드는 문자열 또는 숫자일 수 있습니다.

우편번호 응답률 구하기

메뉴에서 다음을 선택합니다.

다이렉트 마케팅 > 기술 선택

1. **최상위 응답 우편번호 식별**을 선택합니다.
2. 제안에 응답한 연락처를 식별하는 필드를 선택합니다.
3. 긍정 응답을 나타내는 값을 입력합니다. 값에 정의된 값 레이블이 있는 경우, 드롭 다운 목록에서 값 레이블을 선택하면 해당 값이 표시됩니다. 주제를 참조하십시오.
4. 우편번호가 포함된 필드를 선택합니다.
5. 프로시저를 실행하려면 **실행**을 클릭하십시오.

선택적으로 다음을 수행할 수 있습니다.

- 우편번호의 전체 값 대신 처음 n 개의 문자 또는 자리를 기반으로 응답률을 생성합니다.
- 결과를 Excel 파일에 자동으로 저장합니다.
- 출력 표시 옵션을 제어합니다.

설정

지역코드 그룹화 기준

응답률을 계산하기 위해 레코드를 그룹화하는 방법을 결정합니다. 기본적으로 우편번호 전체가 사용되며, 그룹 응답률을 계산하기 위해 우편번호가 동일한 모든 레코드가 함께 그룹화됩니다. 또는 전체 우편번호 중 처음 n 개의 자리 또는 문자로 구성된 일부만을 기준으로 레코드를 그룹화할 수 있습니다. 예를 들어, 10개의 문자로 구성된 우편번호의 처음 5자리 또는 5자리로 구성된 우편번호의 처음 3자리만을 기준으로 레코드를 그룹화할 수 있습니다. 출력 데이터 세트에는 우편번호 그룹마다 하나의 레코드가 포함됩니다. 값을 입력하는 경우 값의 양의 정수여야 합니다.

숫자 우편번호 형식

우편번호 필드가 숫자이고 전체 값 대신 처음 n 자리를 기준으로 우편번호를 그룹화하려는 경우 원래 값의 자릿수를 지정해야 합니다. 자릿수는 우편번호에서 사용 가능한 최대 자릿수입니다. 예를 들어, 우편번호 필드에 5자리 및 9자리 우편번호가 혼재하는 경우 9를 자릿수로 지정해야 합니다.

참고: 표시 형식에 따라 일부 5자리 우편번호는 4자리만 포함하는 것처럼 보일 수 있지만 앞에 0이 내재되어 있습니다.

출력

우편번호별 응답 비율이 있는 새 데이터 세트 외에도 결과를 십분위수 등급(상위 10%, 상위 20% 등)으로 요약하는 테이블 및 차트를 표시할 수 있습니다. 표는 각 십분위의 응답률, 누적 응답률, 레코드 수 및 누적 레코드 수를 표시합니다. 도표는 각 십분위의 누적 응답률 및 누적 레코드 수를 표시합니다.

수용 가능한 최소 응답률. 대상 최소 응답률 또는 손익분기 공식을 입력하는 경우, 테이블에는 최소 누적 응답률을 충족하는 십분위가 색상으로 구별되어 표시되고 도표에는 지정된 최소 응답률 값에서 참조선이 표시됩니다.

- **대상 응답률.** 백분율(각 우편번호 그룹의 긍정 응답 백분율)로 표현된 응답률입니다. 값은 0보다 크고 100보다 작아야 합니다.
- **공식에서 손익분기 비율 계산.** 다음 공식을 기반으로 최소 누적 응답률을 계산합니다. (하나의 패키지 발송 비용/응답당 순이익) x 100. 두 값은 모두 양수여야 합니다. 결과는 0보다 크고 100보다 작아야 합니다. 예를 들어, 패키지 발송 비용이 \$0.75이고 응답당 순이익이 \$56인 경우 최소 응답률은 $(0.75/56) \times 100 = 1.34\%$ 입니다.

연락처 최대 수. 연락처 최대 수를 지정하는 경우, 테이블에는 연락처(레코드)의 누적 최대 수를 초과하지 않는 십분위가 색상으로 구별되어 표시되고 도표에는 해당 값에서 참조선이 표시됩니다.

- **연락처 백분율.** 백분율로 표시되는 최대값입니다. 예를 들어, 모든 연락처 중 50% 이하가 포함되고 응답률이 가장 높은 십분위를 알고자 할 수 있습니다. 값은 0보다 크고 100보다 작아야 합니다.
- **연락처 수.** 연락처 수로 표시되는 최대값입니다. 예를 들어, 발송하는 패키지가 10,000개를 초과하지 않도록 하려는 경우 값을 10000으로 설정할 수 있습니다. 값은 양의 정수여야 합니다(그룹화 기호가 없음).

수용 가능한 최소 응답률과 연락처 최대 수를 둘 다 지정하는 경우 먼저 충족되는 조건에 따라 테이블의 색상이 구별됩니다.

Excel로 내보내기

이 프로시저는 우편번호별 응답률을 포함하는 새 데이터 세트를 자동으로 작성합니다. 데이터 세트의 각 레코드(행)는 우편번호를 나타냅니다. Excel 파일에 동일한 정보를 자동으로 저장할 수 있습니다. 이 파일은 Excel 97-2003 형식으로 저장됩니다.

범주형 응답 필드 작성

응답 필드는 하나의 값이 모든 긍정 응답을 나타내는 범주형 필드여야 합니다. 기타 비결측값은 부정 응답으로 간주됩니다. 응답 필드가 연속형(최도) 값(예: 구매 수 또는 구매 금액)을 나타내는 경우, 0이 아닌 모든 응답값에 단일의 긍정 응답값을 할당하는 새 필드를 작성해야 합니다.

- 부정 응답이 0(공백이 아님. 공백은 결측으로 처리됨)으로 기록되는 경우 이는 다음 공식으로 계산할 수 있습니다.

`newName=OldName>0`

여기서 *NewName*은 새 필드의 이름이고 *OldName*은 원래 필드의 이름입니다. 이것은 0보다 큰 모든 비결측값에 1 값을 할당하고 0 이하의 모든 비결측값에 0을 할당하는 논리식입니다.

- 부정 응답에 대해 아무 값도 기록되지 않는 경우 이러한 값은 결측값으로 처리되며 공식은 좀 더 복잡합니다.

`newName=NOT(MISSING(OldName))`

이 논리식에서는 모든 비결측 응답값에 1 값이 할당되고 모든 결측 응답값에 0 값이 할당됩니다.

- 부정(0) 응답값과 결측값을 구별할 수 없는 경우에는 정확한 응답값을 계산할 수 없습니다. 실제 결측값 수가 비교적 적은 경우에는 계산되는 응답률에 큰 영향을 미치지 않을 수도 있습니다. 그러나 결측값 수가 많은 경우, 예를 들어, 전체 데이터 세트 중 작은 검정용 표본에 대해서만 응답 정보가 기록되는 경우, 계산된 응답률은 실제 응답률보다 현저히 낮기 때문에 의미가 없게 됩니다.

범주형 응답 필드 작성

1. 메뉴에서 다음을 선택합니다.

변환 > 변수 계산

2. 대상 변수에 새 필드(변수) 이름을 입력합니다.
3. 부정 응답이 0으로 기록되는 경우, 수식에 `OldName>0`을 입력합니다. 여기서 *OldName*은 원래 필드 이름입니다.

4. 부정 응답이 결측값(공백)으로 기록되는 경우 수식에 NOT(MISSING(OldName))을 입력합니다. 여기서 OldName은 원래 필드 이름입니다.

구매 성향

구매 성향은 테스트 메일 또는 이전 캠페인의 결과를 사용하여 점수를 생성합니다. 점수는 어느 연락처가 응답할 가능성이 가장 높은지를 나타냅니다. 응답 필드는 테스트 메일 또는 이전 캠페인에 응답한 사람을 나타냅니다. 성향 필드는 비슷한 특성이 있는 연락처가 응답할 확률을 예측하는 데 사용할 수 있는 특성입니다.

이 방법은 이항 로지스틱 회귀를 사용하여 예측 모형을 작성합니다. 예측 모형을 작성하고 적용하는 프로세스는 다음 두 기본 단계로 구성됩니다.

1. 모형을 작성하고 모형 파일을 저장합니다. 해당 결과(종종 **대상**이라고 함)를 알고 있는 데이터 세트를 사용하여 모형을 작성합니다. 예를 들어 다이렉트 메일 캠페인에 반응할 확률이 높은 사람을 예측할 모형을 작성하려는 경우, 반응한 사람과 반응하지 않은 사람에 대한 정보가 이미 들어 있는 데이터 세트로 시작해야 합니다. 예를 들어, 그러한 데이터 세트는 소그룹 고객에 대한 테스트 메일 결과이거나 과거의 비슷한 캠페인에 대한 응답 정보일 수 있습니다.
2. 예측 결과를 얻기 위해 다른 데이터 세트(해당 결과를 모름)에 해당 모형을 적용합니다.

예. 회사의 다이렉트 마케팅 부서는 검정 메일링 결과를 사용하여 나머지 연락처 데이터베이스에 성향 스코어를 할당하고, 다양한 인구 통계학적 특성을 사용하여 반응하고 구매할 것 같은 확률이 높은 연락처를 식별합니다.

출력

이 프로시저는 자동으로 검정 데이터의 성향 점수를 포함하는 새 필드를 데이터 세트에 작성하고 다른 데이터 세트를 스코어링하는 데 사용할 수 있는 XML 모형 파일을 작성합니다. 선택적 진단 출력에는 예측 응답과 실제 응답을 비교하는 전체 모형 품질 지표 및 분류표가 있습니다.

구매 성향 데이터 고려사항

응답 필드. 응답 필드는 문자열 또는 숫자일 수 있습니다. 이 필드에 구매 수 또는 구매 금액을 표시하는 값이 포함되는 경우에는 단일 값이 모든 긍정 응답을 나타내는 새 필드를 작성해야 합니다. 자세한 정보는 [14 페이지의 『범주형 응답 필드 작성』](#) 주제를 참조하십시오.

긍정 응답 값. 긍정 응답 값은 긍정으로 응답한(예를 들어, 구매) 고객을 식별합니다. 기타 모든 비결측 응답 값은 부정 응답을 나타내는 것으로 간주됩니다. 응답 필드에 대해 정의된 값 레이블이 있는 경우 그러한 레이블은 드롭다운 목록에 표시됩니다. 주제를 참조하십시오.

성향 예측 도구. 성향을 예측하는 데 사용되는 필드는 문자열 또는 숫자일 수 있으며 명목형, 순서형 또는 연속형(척도)일 수 있습니다. 그러나 모든 예측 필드에 적절한 측정 수준을 할당하는 것이 중요합니다.

측정 수준. 측정 수준은 결과의 계산에 영향을 미치므로 올바른 측정 수준을 할당하는 것이 중요합니다.

- **명목(Nominal).** 변수의 값이 고유한 순위가 없는 범주(예: 직원이 근무하는 회사의 부서)를 나타내면 변수를 명목으로 취급할 수 있습니다. 종교, 우편번호 또는 종교 단체 등이 명목변수에 해당합니다.
- **순서(Ordinal).** 변수의 값이 일부 고유한 순위(예: 높은 만족도에 대한 높은 만족도에서 서비스 만족도의 레벨)를 나타내는 범주를 나타내는 경우 변수를 서수로 취급할 수 있습니다. 순서변수의 예로는 만족도나 신뢰도를 나타내는 태도 스코어 및 선호도 등급 스코어가 있습니다.
- **연속.** 변수는 값이 의미 있는 메트릭을 사용하여 정렬된 범주를 나타내므로 값 간 거리 비교가 적절할 때 척도(연속형)로 처리될 수 있습니다. 척도변수의 예로는 연령과 수입이 있습니다.

각 필드 옆의 아이콘은 현재 측정 수준을 나타냅니다.

	숫자	문자열	날짜	시간
척도(연속형)		해당사항 없음		
순서				

표 4. 측정 수준 아이콘 (계속)				
	숫자	문자열	날짜	시간
명목				

You can change the measurement level in Variable View of the Data Editor or you can use the Define Variable Properties dialog to suggest an appropriate measurement level for each field.

측정 수준을 알 수 없는 필드

측정 수준 경고는 데이터 세트에서 하나 이상의 변수(필드)에 대해 측정 수준을 알 수 없을 때 표시됩니다. 측정 수준은 이 프로시저의 계산 결과에 영향을 미치기 때문에 모든 변수에 대해 측정 수준을 정의해야 합니다.

데이터 스캔. 활성 데이터 세트의 데이터를 읽고 현재 알 수 없는 측정 수준이 있는 필드에 기본 측정 수준을 할당합니다. 데이터 세트가 큰 경우 시간이 걸릴 수 있습니다.

수동으로 할당. 알 수 없는 측정 수준이 있는 필드를 모두 나열하는 대화 상자를 엽니다. 이 대화 상자에서 해당 필드에 측정 수준을 할당할 수 있습니다. 데이터 편집기의 변수 보기에서도 측정 수준을 할당할 수 있습니다.

이 프로시저에 대해 측정 수준이 중요하기 때문에 모든 필드에 정의된 측정 수준이 있을 때까지는 대화 상자에 액세스하여 이 프로시저를 실행할 수 없습니다.

구매 성향 점수 구하기

메뉴에서 다음을 선택합니다.

다이렉트 마케팅 > 기술 선택

1. **구매 가능성이 가장 큰 연락처 선택**을 선택합니다.
2. 제안에 응답한 연락처를 식별하는 필드를 선택합니다.
3. 긍정 응답을 나타내는 값을 입력합니다. 값에 정의된 값 레이블이 있는 경우, 드롭 다운 목록에서 값 레이블을 선택하면 해당 값이 표시됩니다. 주제를 참조하십시오.
4. 성향을 예측하는 데 사용할 필드를 선택합니다.
다른 데이터 파일을 스코어링할 모형 XML 파일을 저장하려면 다음을 수행하십시오.
5. **XML 파일에 모형 정보 내보내기**를 선택합니다.
6. 디렉토리 경로 및 파일 이름을 입력하거나 **찾아보기**를 클릭하여 모형 XML 파일을 저장할 위치로 이동합니다.
7. 프로시저를 실행하려면 **실행**을 클릭하십시오.
모형 파일을 사용하여 다른 데이터 세트를 스코어링하려면 다음을 수행하십시오.
8. 스코어링 데이터 세트를 엽니다.
9. 스코어링 방법을 사용하여 데이터 세트에 모형을 적용합니다. 메뉴에서 다음을 선택합니다.

유틸리티 > 스코어링 마법사.

설정

모형 검증

모형 검증은 진단 목적으로 훈련 및 검정 그룹을 작성합니다. 진단 출력 섹션에서 분류표를 선택하면 분류표가 비교를 위해 훈련(선택됨) 섹션과 검정(선택 취소됨) 섹션으로 나누어집니다. 모형 검증을 선택하는 경우에는 분류표도 선택해야 합니다. 점수는 훈련 표본에서 생성된 모형을 기반으로 하고 훈련 표본에는 항상 사용 가능한 레코드 총 수보다 적은 수의 레코드가 포함됩니다. 예를 들어, 기본 훈련 표본 크기가 50%일 때, 사용 가능한 레코드의 반만을 기반으로 작성된 모형은 사용 가능한 모든 레코드를 기반으로 작성된 모형만큼 신뢰성이 높지 않을 수 있습니다.

- **훈련 표본 파티션 크기(%).** 훈련 표본에 지정할 레코드의 백분율을 지정하십시오. 응답 필드에 비결측값이 포함된 나머지 레코드가 검정 표본에 할당됩니다. 값은 0보다 크고 100보다 작아야 합니다.
- **시드를 설정하여 결과 복제.** 훈련 및 검정 표본에 레코드가 무작위로 할당되므로, 항상 동일한 시작 난수 시드 값을 지정하지 않는 한 프로시저를 실행할 때마다 서로 다른 결과를 얻을 수 있습니다.

진단 출력

전체 모형 품질. 0에서 1 사이의 값으로 표현되는 전체 모형 품질의 막대형 차트를 표시합니다. 좋은 모형은 0.5보다 큰 값을 가져야 합니다.

분류표. 예측 긍정 및 부정 응답과 실제 긍정 및 부정 응답을 비교하는 테이블이 표시됩니다. 전체적인 정확도를 보면 모형이 얼마나 잘 작동하는지 알 수 있지만 정확하게 예측되는 긍정 응답의 백분율에 더 관심이 있을 수 있습니다.

- **최소 확률.** 점수 값이 지정된 값보다 큰 레코드를 분류표의 예측 긍정 응답 범주에 할당합니다. 프로시저에 의해 생성된 점수는 연락처가 긍정적으로 응답(예: 구매)할 확률을 나타냅니다. 일반적으로 비율로 표시되는 최소 대상 반응률에 가까운 값을 지정해야 합니다. 예를 들어, 5% 이상의 응답률에 관심이 있는 경우 0.05를 지정하십시오. 이 값은 0보다 크고 1보다 작아야 합니다.

코딩변경된 반응 필드의 이름과 레이블

이 프로시저는 응답 필드를 1이 긍정 응답을 나타내고 0이 부정 응답을 나타내는 새 필드로 자동으로 코딩변경하며, 코딩변경된 필드에서 분석이 수행됩니다. 기본 이름 및 레이블을 대체하고 고유의 이름 및 레이블을 제공할 수 있습니다. 이름은 IBM SPSS Statistics 이름 지정 규칙을 따라야 합니다. 주제를 참조하십시오.

점수 저장

성향 점수를 포함하는 새 필드는 원래 데이터 세트에 자동으로 저장됩니다. 점수는 비율로 표시되는 긍정 응답의 확률을 나타냅니다.

- 필드 이름은 IBM SPSS Statistics 이름 지정 규칙을 따라야 합니다. 주제를 참조하십시오.
- 이 필드 이름은 데이터 세트에 이미 있는 필드 이름과 중복되지 않아야 합니다. 동일한 데이터 세트에 이 프로시저를 두 번 이상 실행하는 경우 매번 다른 이름을 지정해야 합니다.

범주형 응답 필드 작성

응답 필드는 하나의 값이 모든 긍정 응답을 나타내는 범주형 필드여야 합니다. 기타 비결측값은 부정 응답으로 간주됩니다. 응답 필드가 연속형(척도) 값(예: 구매 수 또는 구매 금액)을 나타내는 경우, 0이 아닌 모든 응답값에 단일의 긍정 응답값을 할당하는 새 필드를 작성해야 합니다.

- 부정 응답이 0(공백이 아님. 공백은 결측으로 처리됨)으로 기록되는 경우 이는 다음 공식으로 계산할 수 있습니다.

$newName = OldName > 0$

여기서 *NewName*은 새 필드의 이름이고 *OldName*은 원래 필드의 이름입니다. 이것은 0보다 큰 모든 비결측값에 1 값을 할당하고 0 이하의 모든 비결측값에 0을 할당하는 논리식입니다.

- 부정 응답에 대해 아무 값도 기록되지 않는 경우 이러한 값은 결측값으로 처리되며 공식은 좀 더 복잡합니다.

$newName = NOT(MISSING(OldName))$

이 논리식에서는 모든 비결측 응답값에 1 값이 할당되고 모든 결측 응답값에 0 값이 할당됩니다.

- 부정(0) 응답값과 결측값을 구별할 수 없는 경우에는 정확한 응답값을 계산할 수 없습니다. 실제 결측값 수가 비교적 적은 경우에는 계산되는 응답률에 큰 영향을 미치지 않을 수도 있습니다. 그러나 결측값 수가 많은 경우, 예를 들어, 전체 데이터 세트 중 작은 검정용 표본에 대해서만 응답 정보가 기록되는 경우, 계산된 응답률은 실제 응답률보다 현저히 낮기 때문에 의미가 없게 됩니다.

범주형 응답 필드 작성

1. 메뉴에서 다음을 선택합니다.

변환 > 변수 계산

2. 대상 변수에 새 필드(변수) 이름을 입력합니다.
3. 부정 응답이 0으로 기록되는 경우, 수식에 $OldName > 0$ 을 입력합니다. 여기서 *OldName*은 원래 필드 이름입니다.
4. 부정 응답이 결측값(공백)으로 기록되는 경우 수식에 $NOT(MISSING(OldName))$ 을 입력합니다. 여기서 *OldName*은 원래 필드 이름입니다.

패키지 검정 제어

이 방법은 마케팅 캠페인을 비교하여 서로 다른 패키지 또는 제안의 효과에 유의한 차이가 있는지 확인합니다. 캠페인 효과는 응답별로 측정됩니다. 캠페인 필드는 다양한 캠페인(예: 제안 A와 제안 B)을 식별합니다. 응답 필드는 연락처가 캠페인에 응답했는지를 나타냅니다. 응답이 구매 금액(예: "99.99")으로 기록된 경우에는 구매 금액을 선택하십시오. 응답이 단지 연락처가 긍정적으로 응답했는지 여부(예: "예" 또는 "아니오")를 표시하는 경우에는 응답을 선택하십시오.

예. 회사의 다이렉트 마케팅 부서에서 새 패키지 설계가 기존 패키지보다 긍정 응답을 더 많이 생성하는지 확인하려고 합니다. 따라서 새 패키지가 현저히 더 높은 긍정 응답률을 생성하는지 판별하기 위해 테스트 메일을 보냅니다. 테스트 메일은 기존 패키지를 수신하는 제어 그룹과 새 패키지 설계를 수신하는 검정 그룹으로 구성됩니다. 그런 다음 두 그룹의 결과를 비교하여 유의한 차이가 있는지 확인합니다.

출력

출력에는 캠페인 필드로 정의된 각 그룹의 긍정 및 부정 응답 개수 및 백분율을 표시하는 테이블과 서로 현저한 차이가 있는 그룹을 식별하는 테이블이 포함됩니다.

제어 패키지 검정 데이터 고려사항 및 가정

캠페인 필드. 캠페인 필드는 범주형(명목형 또는 순서형)이어야 합니다.

유효성 응답 필드. 유효성 필드에 구매 금액을 선택하는 경우 이 필드는 숫자여야 하며 측정 수준은 연속형(척도)이어야 합니다.

부정(구매 금액의 경우 0 값) 응답값과 결측값을 구별할 수 없는 경우에는 정확한 응답률을 계산할 수 없습니다. 실제 결측값 수가 비교적 적은 경우에는 계산되는 응답률에 큰 영향을 미치지 않을 수도 있습니다. 그러나 결측값 수가 많은 경우, 예를 들어, 전체 데이터 세트 중 작은 검정용 표본에 대해서만 응답 정보가 기록되는 경우, 계산된 응답률은 실제 응답률보다 현저히 낮기 때문에 의미가 없게 됩니다.

가정. 이 프로시저에서는 연락처가 각 캠페인 그룹에 무작위로 할당되었다고 가정합니다. 즉, 특정 인구통계적 특성, 구매 히스토리 또는 기타 특성이 그룹 할당에 영향을 미치지 않으며 모든 연락처가 임의 그룹에 할당될 확률이 동일합니다.

제어 패키지 검정 구하기

메뉴에서 다음을 선택합니다.

다이렉트 마케팅 > 기술 선택

1. **캠페인의 유효성 비교**를 선택합니다.

2. 각 연락처가 속하는 캠페인 그룹(예: 제안 A, 제안 B 등)을 식별하는 필드를 선택합니다. 이 필드는 명목형 또는 순서형이어야 합니다.

3. 응답 유효성을 나타내는 필드를 선택합니다.

응답 필드가 구매 금액인 경우 필드는 숫자여야 합니다.

응답 필드가 단지 연락처가 긍정적으로 응답했는지 여부(예: "예" 또는 "아니오")를 표시하는 경우에는 **응답**을 선택하고 긍정 응답을 나타내는 값을 입력하십시오. 값에 정의된 값 레이블이 있는 경우, 드롭 다운 목록에서 값 레이블을 선택하면 해당 값이 표시됩니다. 주제를 참조하십시오.

1이 긍정 응답을 나타내고 0이 부정 응답을 나타내는 새 필드가 자동으로 작성되고 새 필드에서 분석이 수행됩니다. 기본 이름 및 레이블을 대체하고 고유의 이름 및 레이블을 제공할 수 있습니다. 이름은 IBM SPSS Statistics 이름 지정 규칙을 따라야 합니다. 주제를 참조하십시오.

4. 프로시저를 실행하려면 **실행**을 클릭하십시오.

주의사항

이 정보는 미국에서 제공되는 제품과 서비스를 위해 작성되었습니다. 본 자료는 IBM에서 다른 언어로도 제공될 수 있습니다. 그러나 자료에 접근하기 위해서는 해당 언어로 된 제품 또는 제품 버전의 사본이 필요할 수 있습니다.

IBM은 다른 국가에서 이 책에 기술된 제품, 서비스 또는 기능을 제공하지 않을 수도 있습니다. 현재 사용할 수 있는 제품 및 서비스에 대한 정보는 한국 IBM 담당자에게 문의하십시오. IBM 제품, 프로그램 또는 서비스를 언급했다고 해서 해당 IBM 제품, 프로그램 또는 서비스만을 사용할 수 있다는 것을 의미하지는 않습니다. IBM의 지적 재산권을 침해하지 않는 한, 기능상으로 동등한 제품, 프로그램 또는 서비스를 대신 사용할 수도 있습니다. 그러나 비IBM 제품, 프로그램 또는 서비스의 운영에 대한 평가 및 검증은 사용자의 책임입니다.

IBM은 이 책에서 다루고 있는 특정 내용에 대해 특허를 보유하고 있거나 현재 특허 출원 중일 수 있습니다. 이 책을 제공한다고 해서 특허에 대한 라이선스까지 부여하는 것은 아닙니다. 라이선스에 대한 의문사항은 다음으로 문의하십시오.

07326

서울특별시 영등포구 국제금융로 10, 31FC

한국 아이.비.엠 주식회사한국 아이.비.엠 주식회사KO

2바이트(DBCS) 정보에 관한 라이선스 문의는 한국 IBM에 문의하거나 다음 주소로 서면 문의하시기 바랍니다.

Intellectual Property Licensing

Legal and Intellectual Property Law

IBM Japan Ltd.

19-21, Nihonbashi-Hakozakicho, Chuo-ku

Tokyo 103-8510, Japan

IBM은 타인의 권리 비침해, 상품성 및 특정 목적에의 적합성에 대한 묵시적 보증을 포함하여(단, 이에 한하지 않음) 묵시적이든 명시적이든 어떠한 종류의 보증 없이 이 책을 "현상태대로" 제공합니다. 일부 국가에서는 특정 거래에서 명시적 또는 묵시적 보증의 면책사항을 허용하지 않으므로, 이 사항이 적용되지 않을 수도 있습니다.

이 정보에는 기술적으로 부정확한 내용이나 인쇄상의 오류가 있을 수 있습니다. 이 정보는 주기적으로 변경되며, 변경된 사항은 최신판에 통합됩니다. IBM은 이 책에서 설명한 제품 및/또는 프로그램을 사전 통지 없이 언제든지 개선 및/또는 변경할 수 있습니다.

이 정보에서 언급되는 비IBM 웹 사이트는 단지 편의상 제공된 것으로, 어떤 방식으로든 이들 웹 사이트를 옹호하고자 하는 것은 아닙니다. 해당 웹 사이트의 자료는 본 IBM 제품 자료의 일부가 아니므로 해당 웹 사이트 사용으로 인한 위험은 사용자 본인이 감수해야 합니다.

IBM은 귀하의 권리를 침해하지 않는 범위 내에서 적절하다고 생각하는 방식으로 귀하가 제공한 정보를 사용하거나 배포할 수 있습니다.

(i) 독립적으로 작성된 프로그램과 기타 프로그램(본 프로그램 포함) 간의 정보 교환 및 (ii) 교환된 정보의 상호 이용을 목적으로 본 프로그램에 관한 정보를 얻고자 하는 라이선스 사용자는 다음 주소로 문의하십시오.

07326

서울특별시 영등포구 국제금융로 10, 31FC

서울국제금융센터(31FC)한국 아이.비.엠 주식회사KO

이러한 정보는 해당 조건(예를 들면, 사용료 지불 등)하에서 사용될 수 있습니다.

이 정보에 기술된 라이선스가 부여된 프로그램 및 프로그램에 대해 사용 가능한 모든 라이선스가 부여된 자료는 IBM이 IBM 기본 계약, IBM 프로그램 라이선스 계약(IPLA) 또는 이와 동등한 계약에 따라 제공한 것입니다.

인용된 성능 데이터와 고객 예제는 예시 용도로만 제공됩니다. 실제 성능 결과는 특정 구성과 운영 조건에 따라 다를 수 있습니다.

비IBM 제품에 관한 정보는 해당 제품의 공급업체, 공개 자료 또는 기타 범용 소스로부터 얻은 것입니다. IBM에서는 이러한 제품들을 테스트하지 않았으므로, 비IBM 제품과 관련된 성능의 정확성, 호환성 또는 기타 청구에 대해서는 확신할 수 없습니다. 비IBM 제품의 성능에 대한 의문사항은 해당 제품의 공급업체에 문의하십시오.

IBM의 미래 방향 또는 의도에 관한 언급은 특별한 통지 없이 변경될 수 있습니다.

이 정보에는 일상의 비즈니스 운영에서 사용되는 자료 및 보고서에 대한 예제가 들어 있습니다. 이들 예제에는 개념을 가능한 완벽하게 설명하기 위하여 개인, 회사, 상표 및 제품의 이름이 사용될 수 있습니다. 이들 이름은 모두 가공의 것이며 실제 인물 또는 기업의 이름과 유사하더라도 이는 전적으로 우연입니다.

저작권 라이선스:

이 정보에는 여러 운영 플랫폼에서의 프로그래밍 기법을 보여주는 원어로 된 샘플 응용프로그램이 들어 있습니다. 귀하는 이러한 샘플 프로그램의 작성 기준이 된 운영 플랫폼의 애플리케이션 프로그래밍 인터페이스(API)에 부합하는 애플리케이션을 개발, 사용, 판매 또는 배포할 목적으로 IBM에 추가 비용을 지불하지 않고 이들 샘플 프로그램을 어떠한 형태로든 복사, 수정 및 배포할 수 있습니다. 이러한 샘플 프로그램은 모든 조건하에서 완전히 테스트된 것은 아닙니다. 따라서 IBM은 이러한 프로그램의 신뢰성, 서비스 가능성 또는 기능을 보증하거나 진술하지 않습니다. 본 샘플 프로그램은 일체의 보증 없이 "현상태대로" 제공됩니다. IBM은 귀하의 샘플 프로그램 사용과 관련되는 손해에 대해 책임을 지지 않습니다.

이러한 샘플 프로그램 또는 파생 제품의 각 사본이나 그 일부에는 반드시 다음과 같은 저작권 표시가 포함되어야 합니다.

© Copyright IBM Corp. 2021. 이 코드의 일부는 IBM Corp. 샘플 프로그램에서 파생되었습니다. 샘플 프로그램

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

상표

IBM, IBM 로고 및 ibm.com은 전세계 여러 국가에 등록된 International Business Machines Corp.의 상표 또는 등록상표입니다. 기타 제품 및 서비스 이름은 IBM 또는 타사의 상표입니다. 현재 IBM 상표 목록은 웹 "저작권 및 상표 정보"(www.ibm.com/legal/copytrade.shtml)에 있습니다.

Adobe, Adobe 로고, PostScript 및 PostScript 로고는 미국 및/또는 기타 국가에서 사용되는 Adobe Systems Incorporated의 등록상표 또는 상표입니다.

Intel, Intel 로고, Intel Inside, Intel Inside 로고, Intel Centrino, Intel Centrino 로고, Celeron, Intel Xeon, Intel SpeedStep, Itanium 및 Pentium은 미국 또는 기타 국가에서 사용되는 Intel Corporation 또는 그 계열사의 상표 또는 등록상표입니다.

Linux는 미국 또는 기타 국가에서 사용되는 Linus Torvalds의 등록상표입니다.

Microsoft, Windows, Windows NT 및 Windows 로고는 미국 또는 기타 국가에서 사용되는 Microsoft Corporation의 상표입니다.

UNIX는 미국 및 기타 국가에서 사용되는 The Open Group의 등록상표입니다.

Java 및 모든 Java 기반 상표와 로고는 Oracle 및/또는 그 계열사의 상표 또는 등록상표입니다.

색인

C

cluster 6

R

RFM

고객 데이터 2

구간화 3

트랜잭션 데이터 2

가

구매 성향 12

군집 분석 6

군집 분석(다이렉트 마케팅 옵션) 6

라

로지스틱 회귀(다이렉트 마케팅) 12

아

예상 프로파일(다이렉트 마케팅 옵션) 7

우편번호 응답률 9

자

제어 패키지 검정 15

