

IBM SPSS Categories 22

IBM

참고

이 정보와 이 정보가 지원하는 제품을 사용하기 전에, 57 페이지의 『주의사항』의 정보를 읽으십시오.

제품 정보

이 개정판은 새 개정판에서 별도로 명시하지 않는 한, IBM SPSS Statistics의 버전 22, 릴리스 0. 수정사항 0 및 모든 후속 릴리스와 수정에 적용됩니다.

목차

제 1 장 범주형 데이터에 대한 최적화 척도법 프로시저 소개	1	범위지정	29
최적화 척도법 설명	1	비선형 정준 상관 분석 옵션	29
최적화 척도법 사용 이유	1	OVERALS 명령 추가 기능	30
최적척도 수준 및 측정 수준	2	제 5 장 대응일치 분석	31
최적척도 수준 선택	3	대응일치 분석에서 행 범위지정	32
변환 도표	3	대응일치 분석에서 열 범위지정	32
범주형 코드	3	대응일치 분석 모형	33
응용 분야에 가장 적합한 프로시저 결정	5	대응일치 분석: 통계	34
범주형 회귀분석	6	대응일치 분석 도표	34
범주형 주성분 분석	7	CORRESPONDENCE 명령 추가 기능	35
비선형 정준 상관 분석	7	제 6 장 다중 대응일치 분석	37
대응일치 분석	8	다중 대응일치 분석에서 변수 가중치 정의	38
다중 대응일치 분석	9	다중 대응일치 분석 이산화	38
다차원척도법	10	다중 대응일치 분석 결측값	38
다차원 확장	10	다중 대응일치 분석 옵션	39
최적화 척도법 차트의 가로 세로 비율	11	다중 대응일치 분석 결과	40
제 2 장 범주형 회귀분석(CATREG)	13	다중 대응일치 분석 저장	40
범주형 회귀분석의 척도 정의	14	다중 대응일치 분석 대상도표	41
범주형 회귀분석: 이산화	14	다중 대응일치 분석 변수 도표	41
범주형 회귀분석 결측값	15	MULTIPLE CORRESPONDENCE 명령 추가 기능	41
범주형 회귀분석: 옵션	15	제 7 장 다차원척도법(PROXSCAL)	43
범주형 회귀분석 정규화	16	한 조의 열에 있어서의 근접 행렬	44
범주형 회귀분석: 결과	17	열에서의 근접 행렬	45
범주형 회귀분석: 저장	17	단일 열에서의 근접 행렬	45
범주형 회귀분석 변환 도표	18	데이터로부터 근접 행렬 작성	45
CATREG 명령 추가 기능	18	데이터로부터 척도 계산하기	46
제 3 장 범주형 주성분 분석(CATPCA)	19	다차원척도법 모형 정의	46
CATPCA에서 척도 및 가중치 정의	20	다차원척도법 제한	47
범주형 주성분 분석: 이산화	21	다차원척도법: 옵션	47
범주형 주성분 분석 결측값	21	다차원척도법 도표, 버전 1	48
범주형 주성분 분석: 옵션	22	다차원척도법 도표, 버전 2	49
범주형 주성분 분석 결과	23	다차원척도법 결과	49
범주형 주성분 분석: 저장	23	PROXSCAL 명령 추가 기능	50
범주형 주성분 분석 대상도표	24	제 8 장 다차원 확장(PREFSCAL)	51
범주형 주성분 분석 범주도표	24	다차원 확장법 모형 정의	52
범주형 주성분 분석 로딩 도표	24	다차원 확장법 제한	53
CATPCA 명령 추가 기능	25	다차원 확장법 옵션	53
제 4 장 비선형 정준 상관 분석(OVERALS)	27	다차원 확장법 도표	54
범위 및 척도 정의	28	다차원 확장법 결과	55

PREFSCAL 명령 추가 기능	55	색인.	61
주의사항	57		
상표.	59		

제 1 장 범주형 데이터에 대한 최적화 척도법 프로시저 소개

범주 프로시저에서는 표준 통계 프로시저로는 분석하기 어렵거나 분석이 불가능한 데이터를 분석하는 데 최적화 척도법을 사용합니다. 이 장에서는 각 프로시저를 통해 수행할 수 있는 작업, 각 프로시저에 가장 적합한 상황, 프로시저 간 관계, 이러한 프로시저와 해당 표준 통계 프로시저 사이의 관계에 대해 설명합니다.

참고: 이러한 프로시저와 이 프로시저를 IBM® SPSS® Statistics에서 구현한 결과물은 라이덴 대학교의 사회 행동 과학부 교육 심리학과 소속 연구자들로 이루어진 DTSS(Data Theory Scaling System Group)에서 개발했습니다.

최적화 척도법 설명

최적화 척도법의 기본 개념은 각 변수의 범주에 수량화된 숫자를 할당한 다음, 표준 프로시저를 사용하여 양적 변수에 대한 해법을 구하는 것입니다.

최적척도 값은 사용하는 프로시저의 최적화 기준에 따라 각 변수의 범주에 할당됩니다. 분석의 명목 또는 순서 변수에 대한 원래 레이블과 달리 이러한 척도 값에는 메트릭 특성이 있습니다.

대부분의 범주 프로시저에서 척도화된 각 변수를 최적으로 수량화하는 데는 최소제곱 변경이라는 반복 방법이 사용됩니다. 이 방법에서는 현재 수량화를 사용하여 해법을 찾은 다음 이 해법을 통해 수량화를 업데이트합니다. 업데이트된 수량화를 사용하여 새 해법을 찾고 새로 찾은 해법을 사용하여 수량화를 다시 업데이트하는 과정이 반복되다가 일정 기준에 도달하면 이 과정이 중지됩니다.

최적화 척도법 사용 이유

범주형 데이터는 사회 행동 과학 분야의 조사, 여론 설문조사, 시장 조사 등에서 자주 등장합니다. 사실상 대부분의 조사자들은 거의 전적으로 범주형 데이터만 다룹니다.

대부분의 표준 모형에서도 특별히 범주형 데이터를 분석하기 위한 프로시저를 사용할 수 있지만 다음과 같은 특징을 갖는 데이터에 대한 적용 결과는 대개 만족스럽지 않습니다.

- 관측값이 너무 적은 데이터
- 변수가 너무 많은 데이터
- 변수당 값이 너무 많은 데이터

최적화 척도법을 사용할 때는 범주를 수량화하므로 이러한 상황에서 발생하는 문제를 해결할 수 있습니다. 또한 표준 모형에서 제공하는 특정 기법을 사용할 수 있는 경우라 해도 이 기법을 유용하게 활용할 수 있습니다.

최적화 척도법 결과는 모수 추정값을 해석하지 않고 대개 그래픽 표시를 기반으로 이루어집니다. 최적화 척도법을 사용하면 매우 유용한 예비 분석을 수행할 수 있으므로 다른 IBM SPSS Statistics 모형을 보완할 수 있습니다. 최적화 척도법을 통해 데이터를 시각화하면 조사의 포커스를 좁혀 모형 모수의 해석에 중점을 둔 분석 기반을 형성할 수 있습니다.

최적척도 수준 및 측정 수준

범주 프로시저를 처음 사용하는 경우 이 개념은 매우 혼동될 수 있습니다. 수준을 지정할 때는 변수가 측도되는 수준이 아니라 변수가 척도화되는 수준을 지정해야 합니다. 기본 개념은 수량화할 변수가 그 측도 방법과 관계없이 비선형 관계를 가질 수 있다는 것입니다.

범주 프로시저에 사용되는 측정에는 세 가지 기본 수준이 있습니다.

- 명목 수준에서는 변수의 값이 순서가 지정되지 않은 범주를 나타내는 것으로 간주합니다. 명목 수준이 될 수 있는 변수의 예로는 지역, 우편번호, 종교 및 기타 여러 가지 선택 범주가 있습니다.
- 순서 수준에서는 변수의 값이 순서가 지정된 범주를 나타내는 것으로 간주합니다. 그 예로는 만족도나 신뢰도를 나타내는 태도 척도, 선호도 등급 스코어 등이 있습니다.
- 연속 수준에서는 변수의 값이 범주 사이의 거리 비교가 가능하도록 의미 있는 메트릭에 따라 순서 지정된 범주를 나타내는 것으로 간주합니다. 그 예로는 연령, 수입 등이 있습니다.

예를 들어, 아래 테이블에 나와 있는 것과 같이 지역, 직업 및 연령 변수가 코딩된 경우를 생각해 볼 수 있습니다.

표 1. 지역, 직업 및 연령 코딩 체계

지역 코드	지역 값	작업 코드	작업 값	연령
1	북	1	수습 사원	20
2	남	2	영업 사원	22
3	동	3	관리자	25
4	서			27

이들 값은 각 변수의 범주를 나타냅니다. *region*은 명목변수입니다. *region*에는 네 가지 범주가 있으며 이들 범주에는 고유한 순서가 없습니다. 값 1에서 4까지는 단순히 네 개의 범주를 나타내며 코딩 체계는 전적으로 작성자의 선택에 달린 문제입니다. 한편, 직업은 순서 변수라고 생각할 수 있습니다. 원래 범주는 수습 사원에서 관리자에 이르는 연속적인 단계를 형성합니다. 코드가 클수록 기업에서 더 높은 직위를 나타냅니다. 그러나 여기에서는 순서 정보만 알 수 있을 뿐이며 인접한 범주 사이의 거리에 대해서는 아무것도 알 수 없습니다. 이외는 대조적으로 연령은 연속 변수라고 생각할 수 있습니다. 연령의 경우 각 값 사이의 거리는 그 자체로서 의미를 갖습니다. 20과 22 사이의 거리는 25와 27 사이의 거리와 같지만 22와 25 사이의 거리는 앞의 두 경우보다 그 차이가 더 큼니다.

최적척도 수준 선택

변수에 대해 어떤 최적척도 수준을 지정할지 자동으로 사전 정의하는 고유한 변수 특성은 없습니다. 다만, 의미 있고 해석하기 더 쉬운 방식으로 데이터를 탐색할 수 있습니다. 예를 들어, 연속 수준 변수를 명목 수준에서 분석하는 경우 비선형 변환을 사용하여 차원 수가 더 적은 해법을 만들 수도 있습니다.

다음 두 예제에서는 "명약관화"해 보이는 측정 수준이 가장 좋은 최적척도 수준이 아닐 수도 있음을 보여줍니다. 오브젝트를 연령 그룹으로 정렬하는 변수를 생각해 봅시다. 연령을 연속 변수로 척도화할 수도 있지만 25세 미만의 사람들에게는 연령과 안전도가 정비례하는 반면 60세 이상의 사람들에게는 연령과 안전도가 반비례할 수도 있는 것이 사실입니다. 이 경우 연령을 명목변수로 취급하는 것이 더 좋을 수도 있습니다.

또 다른 예로 정치적 성향에 따라 사람들을 정렬하는 변수를 생각해 볼 수 있습니다. 이 변수는 근본적으로 명목변수라고 판단할 수 있습니다. 그러나 정치적 좌파에서 정치적 우파 순으로 정당의 순서를 지정하는 경우 이 순서를 고려하여 각 정당을 수량화하기 위해 순서 수준의 분석을 사용할 수도 있습니다.

임의의 변수가 반드시 특정 수준이 되어야 한다고 사전 정의하는 변수 특성은 없지만 초보자가 작업하는 데 도움이 되는 몇 가지 일반 지침은 생각해 볼 수 있습니다. 단일 명목 수량화의 경우 일반적으로 범주의 순서를 모르지만 분석을 위해 일정한 순서를 부여해야 하는 경우가 있습니다. 범주의 순서를 알고 있다면 순서 수량화를 시도해 볼 수 있습니다. 범주의 순서를 지정할 수 없다면 다중 명목 수량화를 시도해 볼 수 있습니다.

변환 도표

각 변수를 척도화할 수 있는 수준이 서로 다르면 수량화에 적용되는 제한도 서로 다릅니다. 변환 도표를 사용하면 선택된 최적척도 수준의 결과에 따른 원래 범주와 수량화 사이의 관계를 시각적으로 확인할 수 있습니다. 예를 들어, 변수를 연속 수준으로 취급하는 경우 선형 변환 도표가 작성됩니다. 변수를 순서 수준으로 취급하면 감소하지 않는 변환 도표가 작성됩니다. U 형태(또는 반대 형태)인 명목 수준으로 취급되는 변수에 대한 변환 도표에는 2차 관계가 표시됩니다. 범주 순서를 완전히 바꾸면 명목변수로 명확한 추세가 없는 변환 도표를 생성할 수도 있습니다. 아래 그림에는 변환 도표의 예가 나와 있습니다.

변환 도표는 선택한 최적척도 수준이 어느 정도의 성능을 발휘할지 확인하는 데 특히 적합합니다. 여러 범주를 유사하게 수량화할 수 있는 경우 이들 범주를 하나의 범주로 합칠 수 있습니다. 또는 명목 수준으로 취급된 변수가 증가 추세를 나타내도록 수량화되는 경우 순서 변환으로도 유사한 적합 결과를 얻을 수 있습니다. 이 추세가 선형인 경우 이를 연속 변수로 취급할 수도 있습니다. 그러나 범주를 합치거나 척도 수준을 바꿀 수 있는 경우 분석은 크게 바뀌지 않습니다.

범주형 코드

코딩 체계에 따라서는 원하지 않는 결과가 발생하거나 분석이 불완전하게 이루어질 수 있으므로 범주형 변수를 코딩할 때는 몇 가지 주의가 필요합니다. 다음 표에는 직업에 대해 작성해 볼 수 있는 코딩 체계가 나와 있습니다.

표2. 직업에 대한 대체 코딩 체계

범주	A	B	C	D
수습 사원	1	1	5	1
영업 사원	2	2	6	5

표 2. 직업에 대한 대체 코딩 체계 (계속)

범주	A	B	C	D
관리자	3	7	7	3

일부 범주 프로시저에서는 사용되는 모든 변수 범위를 정의해야 합니다. 이 범위를 벗어난 값은 모두 결측값으로 취급됩니다. 최소 범주 값은 항상 1입니다. 최대 범주 값은 사용자가 지정합니다. 이 값은 변수에 대한 범주의 수가 아니라 가장 큰 범주 값입니다. 예를 들어, 이 테이블에서 도표 A의 최대 범주 값은 3이고 도표 B의 최대 범주 값은 7이지만 두 도표 모두 동일한 세 가지 범주를 코딩합니다.

변수 범위는 분석에서 제외할 범주를 결정합니다. 해당 코드가 지정된 범위를 벗어난 범주는 모두 분석에서 제외됩니다. 이 방법을 사용하면 범주를 간단하게 생략할 수 있지만 의도하지 않은 분석 결과가 나올 수도 있습니다. 최대 범주를 잘못 정의하면 유효한 범주가 분석에서 제외될 수 있습니다. 예를 들어, 도표 B의 경우 최대 범주 값을 3으로 정의하면 1부터 3까지로 코딩된 범주만 직업에 포함되므로 관리자 범주는 결측값으로 취급됩니다. 실제로 3으로 코딩된 범주는 없으므로 분석의 세 번째 범주에는 아무런 케이스도 포함되지 않습니다. 관리자 범주를 모두 생략하려는 경우에는 이 분석이 적절할 수 있습니다. 그러나 관리자를 포함해야 하는 경우라면 최대 범주를 7로 정의하여 7보다 크거나 1보다 작은 값으로 코딩된 값만 결측값으로 취급하도록 해야 합니다.

명목 또는 순서로 취급되는 변수의 경우 범주의 범위는 결과에 영향을 주지 않습니다. 명목변수의 경우 중요한 것은 레이블이며 해당 레이블과 관련된 값은 중요하지 않습니다. 순서 변수의 경우 이를 수량화할 때 범주의 순서가 보존되며 범주 값 자체는 중요하지 않습니다. 범주 순서가 동일한 모든 코딩 체계는 동일한 결과를 갖습니다. 예를 들어, 이 테이블의 처음 세 도표는 직업을 순서 수준에서 분석하는 경우 동일하게 기능합니다. 범주의 순서는 이들 도표에서 모두 같습니다. 반면, 도표 D는 두 번째 및 세 번째 범주와 그 순서가 반대이므로 이들 도표와 다른 결과를 냅니다.

한 변수에 대해 여러 가지 코딩 체계로 동일한 기능을 수행할 수 있지만 가능하면 코드 사이의 차이가 크지 않은 도표를 사용하는 것이 좋습니다. 코드는 프로시저를 통해 생성되는 결과의 양에 영향을 주기 때문입니다. 1과 사용자가 정의한 최대값 사이의 값으로 코딩된 범주는 모두 유효합니다. 이들 범주가 하나라도 비어 있으면 프로시저에 따라 해당 범주에 상응하는 수량화가 결측값 또는 0이 됩니다. 이러한 값 할당은 모두 분석에 영향을 주지 않지만 이들 범주에 대해서도 결과는 생성됩니다. 예를 들어, 도표 B의 경우 직업에는 시스템-결측값이 할당되는 범주 네 개가 포함됩니다. 도표 C의 경우에도 시스템-결측값 지표가 할당되는 범주 네 개가 있습니다. 반면에 도표 A에는 수량화되는 시스템-결측값이 없습니다. 명목 또는 순서로 취급되는 변수에 대해 연속적인 정수를 코드로 사용하면 결과에 영향을 주지 않고도 훨씬 적은 결과를 얻을 수 있습니다.

연속으로 취급되는 변수에 대한 코딩 체계는 순서 케이스보다 더 제한적입니다. 이러한 변수의 경우 연속적인 범주 사이의 차이는 매우 중요합니다. 다음 표에는 연령에 대한 세 가지 코딩 체계가 나와 있습니다.

표 3. 연령에 대한 대체 코딩 체계

범주	A	B	C
20	20	1	1
22	22	3	2
25	25	6	3

표 3. 연령에 대한 대체 코딩 체계 (계속)

범주	A	B	C
27	27	8	4

연속 변수에 대한 코딩을 변경할 때는 항상 범주 사이의 차이를 유지해야 합니다. 이러한 차이를 유지하기 위한 방법 중 하나로 원래 값을 사용할 수도 있습니다. 그러나 이 경우 시스템 결측값 지표가 있는 범주가 여러 개 생성될 수 있습니다. 예를 들어, 도표 A에는 원래 관측값이 사용됩니다. 대응일치 분석을 제외한 모든 범주 프로시저의 경우 최대 범주 값은 27이고 최소 범주 값은 1로 설정되어 있습니다. 처음 19개의 범주는 비어 있고 시스템 결측값 지표가 할당되어 있습니다. 최대 범주 값이 1보다 훨씬 크고 1과 최대값 사이에 빈 범주가 많은 경우 결과를 검토하는 것은 다소 성가신 일이 될 수 있습니다.

결과의 양을 줄이기 위해 코딩변경을 수행할 수도 있습니다. 그러나 연속형 케이스의 경우에는 자동 코딩변경 기능을 사용하지 말아야 합니다. 연속된 정수에 대해 코딩하면 모든 연속 범주 사이의 차이가 1이 되고 그에 따라 모든 수량화의 간격이 동일해집니다. 연속적인 정수로 코딩을 변경하면 변수를 연속형으로 취급할 때 중요하게 간주되던 메트릭 특성이 무효화됩니다. 예를 들어, 이 테이블의 도표 C는 자동으로 코딩변경된 연령에 상응합니다. 범주 22와 25 사이의 차이가 3에서 1로 변경되고 이를 수량화할 때는 두 차이 중 후자가 반영됩니다.

대체 코딩변경 체계에서 범주 사이의 차이를 유지하려면 모든 범주에서 가장 작은 범주 값을 빼고 각 차이에 1을 더해야 합니다. 도표 B는 이러한 변환의 결과입니다. 가장 작은 범주 값인 20을 각 범주에서 빼고 각각의 결과에 1을 더했습니다. 변환된 코드는 최소값이 1이고 각 범주 사이의 차이는 모두 원래 데이터의 경우와 동일합니다. 이제 최대 범주 값은 8이고 처음으로 나오는 0이 아닌 수량화 앞의 0 수량화는 모두 삭제되었습니다. 그러나 도표 B의 결과인 각 범주에 상응하는 0이 아닌 수량화는 도표 A의 수량화와 동일합니다.

응용 분야에 가장 적합한 프로시저 결정

이러한 네 가지 프로시저(대응일치 분석, 다중 대응일치 분석, 범주형 주성분 분석, 비선형 정준 상관 분석)에서 구현하는 기법은 차원 축소라는 다변량 데이터 분석의 일종입니다. 즉, 변수 사이의 관계는 가능하면 두세 개 정도로 적은 차원을 통해 표현됩니다. 이렇게 하면 너무 복잡하거나 자료가 많아 제대로 파악하기 어려운 관계의 구조나 패턴을 좀 더 쉽게 설명할 수 있습니다. 시장 조사 분야에서 이러한 기법은 지각도의 형태를 취할 수 있습니다. 이 프로시저를 사용할 때의 주요 이점은 최적화 척도법의 각기 다른 수준으로 데이터를 수용할 수 있다는 데 있습니다.

범주형 회귀분석에서는 범주형 반응변수와 범주형 예측변수 조합 사이의 관계를 설명합니다. 각 예측변수가 반응변수에 미치는 영향은 상응하는 회귀 가중치를 통해 설명됩니다. 다른 프로시저에서와 마찬가지로 각기 다른 수준의 최적화 척도법을 사용하여 데이터를 분석할 수 있습니다.

다차원척도법 및 다차원 확장법에서는 오브젝트 간 근접성을 사용하여 저차원 공간에 있는 오브젝트 사이의 관계를 설명합니다.

다음은 각 프로시저에 대한 간략한 지침입니다.

- 범주형 독립변수의 조합을 통해 범주형 종속변수의 값을 예측하려면 범주형 회귀분석을 사용합니다.

- 최적적도 수준이 혼합된 단일-변수 세트에서 변동의 패턴을 설명하려면 범주형 주성분 분석을 사용합니다.
- 최적적도 수준이 혼합된 두 개 이상의 변수 세트가 상호 관련될 수 있는 범위를 평가하려면 비선형 정준 상관 분석을 사용합니다.
- 브랜드 선호도 또는 계량 사회학 선택 데이터 등과 같이 이원배치표로 나타낼 수 있는 이차원 분할표 또는 데이터를 분석하려면 대응일치 분석을 사용합니다.
- 모든 변수가 명목 수준에서 분석된다는 더 강한 가정을 하지 않으려는 경우 범주형 다변량 데이터 행렬을 분석하려면 다중 대응일치 분석을 사용합니다.
- 저차원 공간에 있는 단일 오브젝트 세트에 대한 최소제곱 표현을 찾기 위해 근접성 데이터를 분석하려면 다차원척도법을 사용합니다.
- 저차원 공간에 있는 두 오브젝트 세트에 대한 최소제곱 표현을 찾기 위해 근접성 데이터를 분석하려면 다차원 확장법을 사용합니다.

범주형 회귀분석

범주형 회귀분석은 독립(예측변수)변수 세트를 통해 종속(반응)변수를 예측하는 것이 분석 목적인 경우에 사용하는 것이 가장 좋습니다. 다른 모든 최적화 척도법 프로시저와 마찬가지로, 모든 변수의 각 범주에 회귀를 기준으로 최적화된 척도값이 할당됩니다. 범주형 회귀분석을 해법으로 사용하면 변환된 반응과 변환된 예측변수의 가중치를 부여한 조합 사이의 제곱 상관이 최대화됩니다.

다른 범주 프로시저와의 관계. 최적화 척도법이 적용된 범주형 회귀분석은 변수 세트가 두 개이고 그 중 하나에는 종속변수만 포함되어 있는 최적화 척도법 정준상관분석과 비교할 수 있습니다. 후자의 경우 모든 변수 세트 사이의 어딘가에 놓여 있는 알 수 없는 변수와 각 변수 세트를 비교하여 변수 세트의 유사성을 찾아냅니다. 범주형 회귀분석에서는 변환된 예측변수의 선형결합과 변환된 반응의 유사성을 직접 평가합니다.

표준 기법과의 관계. 표준 선형 회귀분석에서는 범주형 변수를 지시변수로 코딩변경하거나 구간 수준 변수와 같은 방식으로 취급할 수 있습니다. 첫 번째 방법의 경우 범주형 변수 수준의 각 조합에 대한 개별 절편과 기울기가 모형에 포함됩니다. 이 경우 많은 수의 모수를 해석해야 합니다. 두 번째 방법에서는 각 변수에 대해 모수를 하나만 추정합니다. 그러나 범주 코딩의 성격이 자의적이므로 일반화는 불가능합니다.

변수 중 일부가 연속적이지 않다면 대체 분석을 사용할 수 있습니다. 반응이 연속적이고 예측변수가 범주형인 경우에는 대개 분산 분석이 사용됩니다. 반응이 범주형이고 예측변수가 연속적인 경우에는 로지스틱 회귀분석이나 판별 분석이 적합할 수 있습니다. 반응과 예측변수가 모두 범주형인 경우에는 로그선형 모형이 자주 사용됩니다.

최적화 척도법에 따른 회귀분석에서는 각 변수에 세 가지 척도 수준을 적용할 수 있습니다. 이들 수준을 조합하면 임의의 "표준" 방법 하나로는 제대로 설명할 수 없는 다양한 비선형 관계를 설명할 수 있습니다. 이와 같이, 최적화 척도법은 표준 방법에 비해 별로 복잡하지 않으면서도 뛰어난 융통성을 제공합니다.

또한 예측변수를 비선형 변환하면 일반적으로 예측변수 사이의 종속성을 줄일 수 있습니다. 예측변수에 대한 상관 행렬의 고유 값을 최적화 척도화된 예측변수에 대한 상관 행렬의 고유 값과 비교해 보면 전자에 비해 후자의 변수 세트에는 일반적으로 변수가 더 적다는 사실을 알 수 있습니다. 즉, 범주형 회귀분석에서 최적화 척도법을 사용하면 예측변수 상관 행렬의 큰 고유 값은 더 작아지고 작은 고유 값은 더 커집니다.

범주형 주성분 분석

범주형 주성분 분석은 최적적도 수준이 혼합된 단일-변수 세트에서 변동의 패턴을 설명하는 경우 가장 적합한 방법입니다. 이 기법을 사용하면 변수 세트의 차원을 줄이면서 최대한 많은 변동을 설명할 수 있습니다. 모든 변수의 각 범주에 척도값이 할당되고 주성분 해법을 기준으로 이러한 값이 최적화됩니다. 분석할 오브젝트에는 수량화된 데이터를 기반으로 성분 스코어가 부여됩니다. 성분 스코어 도표를 살펴보면 분석 오브젝트 간의 패턴을 파악할 수 있고 데이터에서 특수 오브젝트를 찾아낼 수 있습니다. 범주형 주성분 분석을 해법으로 사용하면 지정된 성분(차원) 수에 대해 각 양적변수와 오브젝트 스코어의 상관을 최대화할 수 있습니다.

범주형 주성분 분석을 적용해야 하는 경우의 중요한 예로는 반응자가 선호도에 따라 여러 가지 항목의 순위 또는 등급을 매기는 선호도 데이터 조사를 들 수 있습니다. 일반적인 IBM SPSS Statistics 데이터 설정에서 행은 개인을 나타내고 열은 항목에 대한 측정값을 나타내며 행 간 스코어는 데이터를 행 조건적으로 만드는 선호도 스코어(예를 들어, 0에서 10까지의 척도)를 나타냅니다. 선호도 데이터와 관련하여 개인을 변수로 취급할 수도 있습니다. 전치 프로시저를 사용하면 데이터를 전치할 수 있습니다. 평가자는 변수가 되고 모든 변수는 순서로 선언됩니다. CATPCA에서는 오브젝트보다 변수를 더 많이 사용할 수 있습니다.

다른 범주 프로시저와의 관계. 모든 변수가 다중 명목변수로 선언된 경우 범주형 주성분 분석을 실행하면 동일한 변수에 대해 다중 대응일치 분석을 실행할 때와 같은 분석이 이루어집니다. 따라서 범주형 주성분 분석은 변수 중 일부가 순서 또는 연속형으로 선언된 다중 대응일치 분석의 한 유형이라고 이해할 수도 있습니다.

표준 기법과의 관계. 모든 변수가 연속 수준에서 척도화된 경우 범주형 주성분 분석은 표준 주성분분석과 같습니다.

더 일반적으로 말해 범주형 주성분 분석은 비연속 척도 사이의 상관을 계산하고 표준 주성분 또는 요인 분석 방법을 사용하여 이를 분석하는 대신 사용할 수 있는 분석 방법입니다. 일반적인 Pearson 상관계수를 그대로 순서 데이터의 연관 측도로 사용하면 상관에 대한 추정이 크게 편향될 수 있습니다.

비선형 정준 상관 분석

비선형 정준 상관 분석은 여러 가지 분야에 적용되는 매우 일반적인 프로시저입니다. 비선형 정준 상관 분석의 목적은 주성분분석에서와 같이 변수 자체 사이의 관계를 분석하는 것이 아니라 둘 이상의 변수 세트 사이의 관계를 분석하는 데 있습니다. 예를 들어, 두 개의 변수 세트가 있고 그 중 한 변수 세트는 반응자 세트에 대한 인구 통계학적 배경 항목을 포함하며 다른 한 변수 세트는 태도 항목 세트에 대한 반응을 포함할 수 있습니다. 분석의 척도 수준으로는 명목, 순서 및 연속 수준을 임의로 혼합하여 사용할 수 있습니다. 최적화 척도법 정준상관분석에서는 각 변수 세트의 정준 변수를 오브젝트에 할당된 스코어의 상관 세트와 동시에 비교하여 변수 세트 사이의 유사성을 확인합니다.

다른 범주 프로시저와의 관계. 두 개 이상의 변수 세트가 있고 각 변수 세트마다 변수가 하나씩만 포함된 경우 최적화 척도법 정준상관분석은 최적화 척도법 주성분분석과 같습니다. 변수 세트별로 변수가 하나씩인 분석의 모든 변수가 다중 명목인 경우 최적화 척도법 정준상관분석은 다중 대응일치 분석과 같습니다. 변수 세트가 두 개이고 그 중 하나에는 변수가 하나만 포함되어 있는 경우 최적화 척도법 정준상관분석은 최적화 척도법의 범주형 회귀분석과 같습니다.

표준 기법과의 관계. 표준 정준상관분석은 한 변수 세트의 선형결합과 다른 변수 세트의 선형결합에서 최대로 상호 관련된 선형결합을 찾아내는 통계 기법입니다. 정준상관분석에서는 이러한 선형결합 변수 세트를 찾는 다음, 후속적인 독립 선형결합 변수 세트를 찾아냅니다. 이를 정준 변수라고 하며 최대 수가 더 작은 변수 세트의 변수 수에 도달할 때까지 이 작업은 계속됩니다.

분석할 변수 세트가 두 개이고 모든 변수가 연속형으로 정의된 경우 최적화 척도법 정준상관분석은 표준 정준상관분석과 같습니다. IBM SPSS Statistics에는 정준상관분석 프로시저가 없지만 분산에 대한 다변량 분석을 통해 여러 가지 관련 통계를 구할 수 있습니다.

최적화 척도법 정준상관분석을 적용할 수 있는 분야에는 여러 가지가 있습니다. 두 개의 변수 세트가 있고 그 중 하나에 단일 명목으로 선언된 명목변수가 포함되어 있는 경우 최적화 척도법 정준상관분석 결과는 회귀분석의 경우와 유사한 방식으로 해석할 수 있습니다. 변수를 다중 명목으로 간주하는 경우 판별 분석 대신 최적화 척도법 분석을 사용할 수 있습니다. 변수를 두 개 이상의 변수 세트로 그룹화하면 데이터를 여러 가지 방법으로 분석할 수 있습니다.

대응일치 분석

대응일치 분석은 대응일치표에 대한 Bi-플롯을 만드는 것을 목적으로 합니다. 대응일치표에서 행 및 열 변수는 순서가 지정되지 않은 범주를 나타내는 것으로 간주됩니다. 따라서 항상 명목 최적척도 수준이 사용됩니다. 두 변수는 모두 해당 명목 정보만 검사됩니다. 즉, 일부 오브젝트는 동일한 범주에 속하는 반면 다른 오브젝트는 그렇지 않다는 사실만 고려됩니다. 동일한 변수의 범주 사이에서 거리나 순서에 대한 내용은 아무것도 가정하지 않습니다.

대응일치 분석의 특별한 용례 중 하나로 이원분할표에 대한 분석을 들 수 있습니다. 테이블에 r 개의 활성 행과 c 개의 활성 열이 있는 경우 대응일치 분석 해법의 차원 수는 r 에서 1을 뺀 값이나 c 에서 1을 뺀 값 중 작은 쪽의 최소값입니다. 즉, 분할표의 행 범주 또는 열 범주를 차원의 공간으로 완벽하게 나타낼 수 있습니다. 그러나 2차원 도표는 다차원 공간 표현보다 더 쉽게 이해할 수 있으므로 실제로 있어서는 이원표의 행 및 열 범주를 저차원 공간, 즉 2차원으로 표현하는 경우가 많습니다.

가능한 차원의 최대 수보다 적게 사용하는 경우, 분석을 통해 생성된 통계로 행 및 열 범주가 저차원 표현에서 얼마나 잘 표현되는지 알 수 있습니다. 2차원 해법의 표현 수준이 우수하다면 행 점과 열 점의 도표를 조사하여 행 변수의 어떤 범주가 유사하고 열 변수의 어떤 범주가 유사한지 확인할 수 있고 행 및 열의 어떤 범주가 서로 유사한지 확인할 수 있습니다.

다른 범주 프로시저와의 관계. 단순 대응일치 분석은 이원표로 제한됩니다. 필요한 변수가 세 개 이상이라면 변수를 결합하여 상호작용 변수를 만들 수 있습니다. 예를 들어, 지역, 직업 및 연령 변수에서 지역 및 직업을 결합하여 다음 테이블에서와 같이 12개의 범주가 포함된 지역과직업이라는 새 변수를 만들 수 있습니다. 이 새로운 변수는 대응일치 분석을 통해 분석할 수 있는 연령(12행, 4열)이 포함된 이원표를 형성합니다.

표 4. 지역 및 직업 결합

범주 코드	범주 정의	범주 코드	범주 정의
1	북부, 수습 사원	7	동부, 수습 사원
2	북부, 영입 사원	8	동부, 영입 사원

표4. 지역 및 직업 결합 (계속)

범주 코드	범주 정의	범주 코드	범주 정의
3	북부, 관리자	9	동부, 관리자
4	남부, 수습 사원	10	서부, 수습 사원
5	남부, 영업 사원	11	서부, 영업 사원
6	남부, 관리자	12	서부, 관리자

이 방법의 한 가지 단점은 임의의 변수 쌍이 결합될 수 있다는 점입니다. 직업 및 연령을 결합하여 12개의 범주로 구성된 다른 변수를 만들 수 있습니다. 또는 지역 및 연령을 결합하여 16개의 범주로 구성된 새로운 변수를 만들 수도 있습니다. 이러한 상호작용 변수는 각각 나머지 변수와의 이원표를 형성합니다. 이들 세 가지 테이블에 대한 대응일치 분석 결과는 서로 다르지만 그 각각은 유효한 접근 방법입니다. 또한 변수가 네 개 이상이라면 상호작용 변수를 다른 상호작용 변수와 비교하는 이원표가 작성될 수도 있습니다. 변수가 그리 많지 않더라도 분석 가능한 테이블의 수는 상당히 많아질 수 있습니다. 이러한 테이블 중 하나를 선택하여 분석하거나 테이블 전체를 분석할 수 있습니다. 또는 다중 대응일치 분석 프로시저를 사용하여 상호작용 변수를 작성할 필요 없이 모든 변수를 동시에 조사할 수 있습니다.

표준 기법과의 관계. 교차 분석표 프로시저를 사용하여 분할표를 분석할 수도 있습니다. 이때 분석의 공통 초점은 독립성에 주어집니다. 그러나 테이블의 크기가 작더라도 독립성에서의 이탈 요인을 발견하기는 어려울 수 있습니다. 대응일치 분석의 이점은 크기에 상관없이 이원표에 대해 독립성 이탈 패턴을 표시할 수 있다는 것입니다. 행 변수와 열 변수 사이에 연관이 있는 경우, 즉 카이제곱 값이 유의한 경우 대응일치 분석을 사용하면 관계의 본질을 밝혀내는 데 도움이 될 수 있습니다.

다중 대응일치 분석

다중 대응일치 분석에서는 동일한 범주 내의 오브젝트가 근접하여 함께 도표화되고 서로 다른 범주의 오브젝트는 멀리 떨어져 도표화되는 해법을 작성합니다. 각 오브젝트는 오브젝트에 적용되는 범주의 범주 점에 가능한 한 가깝게 배치됩니다. 이와 같은 방식을 통해 범주는 오브젝트를 동종 하위 그룹으로 분류합니다. 변수는 동일한 범주의 오브젝트를 동일한 하위 그룹으로 분류하는 경우 동일군으로 간주됩니다.

일차원 해법의 경우 다중 대응일치 분석에서는 평균적으로 전체 범주의 산포도가 최대가 되도록 각 변수의 각 범주에 최적척도 값(범주 수량화)을 할당합니다. 이차원 해법의 경우 다중 대응일치 분석에서는 첫 번째 세트와 관련되지 않은 각 변수의 범주에 대한 두 번째 수량화 세트를 찾아 해당 산포도를 다시 최대화하는 과정을 수행합니다. 변수의 범주에는 차원만큼 다양한 스코어가 부여되므로 이 분석의 변수는 최적척도 수준에서 다중 명목으로 간주됩니다.

또한 다중 대응일치 분석에서는 범주 수량화가 해당 범주의 오브젝트에 대한 오브젝트 스코어의 평균 또는 중심값이 되도록 분석 오브젝트에 스코어를 할당합니다.

다른 범주 프로시저와의 관계. 다중 대응일치 분석을 동질성 분석 또는 이중 척도화라고도 합니다. 변수가 두 개만 있는 경우 이 분석을 실행하면 대응일치 분석의 경우와 비견할 수 있지만 동일하지는 않은 결과가 생성됩니다. 대응일치 분석에서는 안정성 정보를 비롯하여 해법의 표현에 대한 적합 및 수준을 요약하는 고유한 결과가 생성됩니다. 따라서 케이스의 변수가 두 개인 경우에는 다중 대응일치 분석보다 대응일치 분석을 사용하는

것이 좋습니다. 두 프로시저 간의 또 다른 차이를 들면, 다중 대응일치 분석의 입력은 행이 오브젝트를 나타내고 열이 변수를 나타내는 데이터 행렬인 반면, 대응일치 분석의 입력은 이와 동일한 데이터 행렬, 일반 근접 행렬 또는 행과 열이 모두 변수의 범주를 나타내는 통합 행렬인 결합 분할표가 될 수 있다는 점입니다. 다중 대응일치 분석은 다중 명목수준에서 척도화된 데이터에 대한 주성분분석이라고 생각할 수도 있습니다.

표준 기법과의 관계. 다중 대응일치 분석은 다원분할표를 대상으로 하는 분석이라고 생각할 수 있습니다. 다원 분할표는 교차 분석표 프로시저로 분석할 수도 있지만 교차 분석표의 경우 각 제어변수의 각 범주에 대해 개별 요약 통계가 생성됩니다. 다중 대응일치 분석의 경우 대개 모든 변수와 단일 2차원 도표 사이의 관계를 요약할 수 있습니다. 다중 대응일치 분석을 응용하면 원래 범주 값을 첫 번째 차원의 최적척도 값으로 바꾸고 두 번째 다변량 분석을 수행할 수 있습니다. 다중 대응일치 분석에서는 범주 레이블 대신 숫자 척도 값을 사용하므로 다중 대응일치 분석을 수행한 후 숫자 데이터가 필요한 다양한 프로시저를 적용할 수 있습니다. 예를 들어, 요인 분석 프로시저를 수행하여 다중 대응일치 분석의 첫 번째 차원에 해당하는 첫 번째 주성분을 생성할 수 있습니다. 첫 번째 차원의 성분 스코어는 오브젝트 스코어와 같고, 제공한 성분 로딩은 판별 척도와 같습니다. 그러나 두 번째 다중 대응일치 분석 차원은 요인 분석의 두 번째 차원과 같지 않습니다.

다차원척도법

다차원척도법은 오브젝트 또는 케이스의 단일 세트 사이에서 거리척도 세트의 구조를 찾아내는 것이 분석 목적인 경우에 사용하는 것이 가장 좋습니다. 이를 위해 개념적 저차원 공간에 있는 점 사이의 거리가 주어진 유사성(상이성) 척도와 가능한 한 가깝게 일치되도록 이 공간의 특정 위치에 관측값이 할당됩니다. 그 결과 해당 저차원 공간에 오브젝트의 최소제곱 표현이 얻어지며, 이 표현을 통해 데이터를 더 심도 있게 이해할 수 있습니다.

다른 범주 프로시저와의 관계. 다변량 데이터를 사용하여 거리를 작성한 다음, 다차원 척도법으로 분석하는 경우 그 결과는 오브젝트 주성분 정규화와 함께 범주형 주성분 분석을 사용하여 데이터를 분석하는 경우와 비슷합니다. 이 유형의 PCA를 주좌표분석이라고도 합니다.

표준 기법과의 관계. 범주 다차원척도법 프로시저(PROXSCAL)를 사용하면 Statistics Base 옵션에서 사용할 수 있는 척도법 프로시저(ALSCAL)를 사용할 때에 비해 여러 가지 이점을 얻을 수 있습니다. PROXSCAL은 특정 모형에 대해 속도가 향상된 알고리즘을 제공하며 이를 통해 공통 공간에 제약을 둘 수 있습니다. 또한 PROXSCAL에서는 긴장이라고도 하는 S-스트레스 대신 정규화된 원시 스트레스를 최소화하는 시도도 합니다. 정규화된 원시 스트레스는 거리를 기반으로 한 척도인 반면 S-스트레스는 제공거리를 기반으로 하므로 일반적으로 정규화된 원시 스트레스를 사용하는 것이 좋습니다.

다차원 확장

다차원 확장법은 행 및 열 오브젝트라고 하는 두 개의 오브젝트 세트 사이에서 거리 척도 세트의 구조를 찾아내는 것이 분석 목적인 경우 적합한 방법입니다. 이를 위해 개념적 저차원 공간에 있는 점 사이의 거리가 주어진 유사성(상이성) 척도와 가능한 한 가깝게 일치되도록 이 공간의 특정 위치에 관측값이 할당됩니다. 그 결과는 해당 저차원 공간에 있는 행 및 열 오브젝트의 최소제곱 표현입니다. 대부분의 경우 이 표현은 데이터를 더 심도 있게 이해하는 데 도움이 됩니다.

다른 범주 프로시저와의 관계. 데이터가 단일 오브젝트 세트 사이의 거리로 이루어진 경우(제공 대칭 행렬) 다차원척도법을 사용합니다.

표준 기법과의 관계. 범주 다차원 확장법 프로시저(PREFSCAL)를 사용하면 (ALSCAL을 통해) Statistics Base 옵션에서 제공하는 확장 기능을 사용할 때에 비해 여러 가지 이점을 얻을 수 있습니다. PREFSCAL을 사용하면 공통 공간에 제약을 둘 수 있습니다. 또한 PREFSCAL에서는 이전의 알고리즘에서 흔히 발생하는 부적절한 해법을 방지할 수 있도록 페널티가 부여된 스트레스 측도를 최소화하는 시도를 합니다.

최적화 척도법 차트의 가로 세로 비율

최적화 척도법 도표의 가로 세로 비율은 등방성입니다. 2차원 도표에서 차원 1의 한 단위를 나타내는 거리는 차원 2의 한 단위를 나타내는 거리와 같습니다. 2차원 도표에서 차원의 범위를 변경하면 실제 거리를 동일하게 유지할 수 있도록 다른 차원의 크기도 자동으로 변경됩니다. 최적화 척도법 프로시저의 경우 등방성 가로 세로 비율을 바꿀 수 없습니다.

제 2 장 범주형 회귀분석(CATREG)

범주형 회귀분석에서는 범주에 숫자 값을 할당하여 범주형 데이터를 수량화함으로써 변환된 변수에 대한 최적 선형 회귀 방정식을 작성합니다. 범주형 회귀분석은 약자로 CATREG(CATegorical REGression)라고도 합니다.

표준 선형 회귀분석은 반응(종속) 변수와 예측변수(독립) 기중 결합의 차이 제곱값 합계를 최소화하는 것입니다. 변수는 보통 양적변수이고 (명목) 범주형 데이터를 사용하는 변수는 이분형이나 대비 변수로 코딩변경됩니다. 이에 따라 범주형 변수는 개별 케이스 그룹으로 제시되고 이 분석 기법을 사용하여 각 그룹의 개별 모수 세트를 추정합니다. 추정된 계수는 예측변수의 변경이 반응에 미치는 영향을 반영합니다. 예측변수 값의 모든 조합에 대하여 반응 예측이 가능합니다.

또 다른 방법은 범주형 예측변수 값 자체에 대한 반응을 회귀분석하는 것입니다. 즉, 각 변수에 대해 하나의 계수를 추정하는 것입니다. 그러나 범주형 변수의 경우 범주값이 임의적이기 때문에 여러 방법으로 범주를 코딩하면 다른 계수가 만들어져 같은 변수의 비교 분석이 어렵게 됩니다.

범주형 회귀분석(CATREG)에서는 명목, 순서 및 연속 변수를 동시에 척도화함으로써 표준 회귀분석을 확장시킵니다. 이 프로시저에서는 범주형 변수를 수량화하여 원래 범주의 특성을 반영합니다. 이 프로시저에서는 수량화된 범주형 변수를 연속 변수와 동일한 방법으로 처리합니다. 비선형 변환을 사용하면 변수를 여러 다양한 수준으로 분석하여 최적의 적합 모형을 찾을 수 있습니다.

예제. 범주형 회귀분석을 사용하여 직업 범주에 따른 직업 만족도, 지역, 여행 횟수 등을 설명할 수 있습니다. 높은 만족도는 관리자 및 낮은 여행 빈도와 대응되는 것을 알 수 있습니다. 회귀 방정식 결과를 통해 위의 세 가지 독립변수 조합에 대한 직업 만족도를 예측할 수 있습니다.

통계 및 도표. 빈도, 회귀계수, 분산 분석표, 반복 히스토리, 범주 수량화, 변환되지 않은 예측변수 간의 상관, 변환된 예측변수 간의 상관, 잔차 도표, 변환 도표 등을 포함합니다.

범주형 회귀분석 데이터 고려 사항

데이터. 범주형 회귀분석은 양의 정수인 범주 표시기에 대해서만 수행됩니다. 이산화 대화 상자를 사용하면 분수값 변수와 문자열 변수를 양의 정수로 변환할 수 있습니다.

가정. 반응변수는 하나만 사용할 수 있지만 예측변수는 최대 200개까지 사용할 수 있습니다. 데이터에는 셋 이상의 유효 케이스가 있어야 하고 유효 케이스의 수는 예측변수의 수에 1을 더한 수보다 많아야 합니다.

관련 프로시저. CATREG는 관련된 범주 세트가 둘이고 그 중 하나에 변수가 하나 있는 경우 최적화 척도법을 사용하는 범주형 정준상관분석(OVERALS)과 동일합니다. 모든 변수를 연속적으로 척도화하는 것은 표준 다중 회귀분석에 해당됩니다.

범주형 회귀분석 수행 방법

1. 메뉴에서 다음을 선택합니다.

분석 > 회귀분석 > 최적화 척도법(CATREG)...

2. 종속변수와 독립변수를 선택합니다.

3. 확인을 클릭합니다.

선택적으로 각 변수에 대한 척도 수준을 바꿉니다.

범주형 회귀분석의 척도 정의

종속변수와 독립변수에 대한 최적척도 수준을 설정할 수 있습니다. 기본적으로 두 개의 내부 마디가 있는 2차 단조 스플라인(순서)으로 척도화됩니다. 추가적으로 분석변수의 가중치를 설정할 수 있습니다.

최적척도 수준. 각 변수를 수량화하는데 사용하는 척도 수준을 선택할 수도 있습니다.

- 스플라인 순서. 관측 변수에 대한 범주 순서는 최적척도화 변수에 보관됩니다. 범주점은 원점을 지나는 직선(벡터)이 됩니다. 변환 결과는 선택한 차수의 평활 단조 조각별 다항입니다. 조각은 내부 마디에 대한 사용자 정의 수와 프로시저 지정 배치에 따라 지정됩니다.
- 스플라인 명목. 관측범주의 범주 오브젝트 그룹만 유지하면서 변수를 최적수준으로 척도화합니다. 관측변수의 범주 순서는 유지되지 않습니다. 범주점은 원점을 지나는 직선(벡터)이 됩니다. 변환 결과는 평활이며 단조가 아닐 수도 있는 선택한 차수의 조각별 다항입니다. 조각은 내부 마디에 대한 사용자 정의 수와 프로시저 지정 배치에 따라 지정됩니다.
- 순서. 관측 변수에 대한 범주 순서는 최적척도화 변수에 보관됩니다. 범주점은 원점을 지나는 직선(벡터)이 됩니다. 변환 결과는 스플라인 순서 변환보다 더 잘 적합되지만 평활 정도가 약합니다.
- 명목. 관측범주의 범주 오브젝트 그룹만 유지하면서 변수를 최적수준으로 척도화합니다. 관측변수의 범주 순서는 유지되지 않습니다. 범주점은 원점을 지나는 직선(벡터)이 됩니다. 변환 결과는 스플라인 명목 변환보다 더 잘 적합되지만 평활의 정도가 약합니다.
- 숫자. 범주는 동일한 간격(구간 수준)으로 순서화된 것으로 간주됩니다. 범주 순서를 유지하고 관측변수 간 범주를 동일한 간격으로 유지하면서 최적수준으로 척도화합니다. 범주점은 원점을 지나는 직선(벡터)이 됩니다. 모든 변수가 숫자 수준일 경우 분석은 표준 주성분 분석과 유사합니다.

범주형 회귀분석: 이산화

이산화 대화 상자에서 변수의 코딩을 변경하는 방법을 선택할 수 있습니다. 분수값 변수는 일곱 개의 범주로 (또는 이 숫자가 7 미만인 경우 개별값의 수로) 그룹화되는데 이 범주는 특별히 지정하지 않는 한 근사적으로 정규 분포를 따릅니다. 문자열 변수를 문자순서의 오름차순에 따라 범주 표시기를 할당하여 양의 정수로 변환합니다. 문자열 변수의 이산화는 이러한 정수에 적용됩니다. 기본적으로 다른 변수는 변하지 않습니다. 그 다음 이산화 변수를 분석에 사용합니다.

방법. 그룹, 순위, 곱하기 중에서 선택합니다.

- 그룹. 지정한 범주의 수로 코딩을 변경하거나 구간별로 코딩을 변경합니다.

- **순위.** 케이스의 순위를 정하여 변수를 이산화합니다.
- **곱하기.** 변수의 현재 값을 표준화하고 10을 곱하고, 반올림하고, 가장 작은 이산화 값이 1이 되도록 상수 값을 더합니다.

그룹. 그룹으로 변수를 이산화할 경우 다음 옵션을 사용할 수 있습니다.

- **범주의 수.** 범주의 수를 지정하고 변수값이 이러한 범주에 따라서 근사적으로 정규 분포 또는 균일분포를 따라야 하는지 지정합니다.
- **동일 구간.** 변수를 동일한 크기의 구간으로 정의된 범주로 코딩변경합니다. 구간의 길이를 지정해야 합니다.

범주형 회귀분석 결측값

결측값 대화 상자에서는 분석변수와 보조변수에 있는 결측값 처리에 대한 처리 방법을 선택할 수 있습니다.

처리 방법. 결측값이 있는 오브젝트를 제외(목록별 삭제)하거나 결측값을 대치(능동 처리)하도록 선택합니다.

- **해당 변수에 결측값이 있는 오브젝트 제외** 선택한 변수에 결측값이 있는 오브젝트는 분석에서 제외됩니다. 이 처리 방법은 보조변수에서는 사용할 수 없습니다.
- **결측값 대치** 선택한 변수에서 결측값을 가진 오브젝트에는 대치된 값이 있습니다. 대치법을 선택할 수 있습니다. 결측값을 가장 빈도가 높은 범주로 대체하려면 **최빈값**을 선택합니다. 여러 개의 최빈값이 있는 경우에는 가장 작은 범주 표시기가 있는 값을 사용합니다. 결측값을 임시범주의 동일한 수량화로 대체하려면 **임시범주**를 선택합니다. 이 변수에서 결측값이 있는 오브젝트는 동일한(임시) 범주에 속하는 것으로 간주됩니다.

범주형 회귀분석: 옵션

옵션 대화 상자에서 초기설정 유형 선택, 반복 및 수렴 기준 지정, 보조 오브젝트 선택, 도표 레이블 지정 등의 작업을 수행할 수 있습니다.

보조 오브젝트. 이 옵션을 사용하여 보조 오브젝트로 처리할 오브젝트를 지정할 수 있습니다. 보조 오브젝트 수를 입력하거나 케이스 범위를 지정하고 추가를 클릭하면 됩니다. 보조 오브젝트에 가중치를 줄 수 없습니다. 지정한 가중치는 무시됩니다.

초기설정. 변수가 명목 수준으로 처리되지 않은 경우 연속 설정을 선택합니다. 하나 이상의 변수가 명목 수준으로 처리되는 경우 임의 설정을 선택합니다.

또는 하나 이상의 변수가 순서 또는 순서 스플라인 척도 수준을 가진 경우 일반적인 모형 적합 알고리즘은 차선의 솔루션이 될 수 있습니다. 검정할 모든 부호 패턴을 사용하는 **다중 계통 시작**을 선택하면 항상 최적의 솔루션을 찾지만 데이터 세트의 순서 및 스플라인 순서 변수가 증가함에 따라 필요한 진행 시간이 빠르게 증가합니다. 분산 임계값의 손실율(%)을 지정하여 검정 패턴 수를 줄일 수 있습니다. 여기서 임계값이 증가할 수록 더 많은 부호 패턴이 제외됩니다. 이 옵션을 사용하면 최적의 솔루션을 얻지 못할 수도 있으며 차선의 솔루션을 얻을 확률이 줄어들게 됩니다. 또한 최적의 솔루션을 찾을 수 없는 경우 차선의 솔루션이 최적의 솔루션과

매우 다를 수 있는 확률이 줄어들게 됩니다. 다중 계통 시작이 요청된 경우 각 시작에 대한 회귀계수 부호는 현재 세션의 외부 IBM SPSS Statistics 데이터 파일이나 데이터 세트에 기록됩니다. 자세한 정보는 17 페이지의 『범주형 회귀분석: 저장』 주제를 참조하십시오.

다중 계통 시작을 사용한 이전 실행 결과에는 회귀계수에 대한 고정 부호를 사용할 수 있습니다. 1과 -1로 표시되는 부호는 지정된 데이터 세트나 파일의 행에 있어야 합니다. 정수값인 시작 번호는 사용할 부호가 포함된 이 파일에서 행의 케이스 번호가 됩니다.

기준. 반복을 수행하는 회귀분석의 최대반복수를 지정할 수 있습니다. 또한 수렴 기준 값을 선택할 수 있습니다. 회귀분석에서는 마지막 두 번의 반복 간의 전체 적합 차이가 수렴값 미만이거나 최대반복수에 도달한 경우에 분석이 중단됩니다.

도표 레이블기준변수. 도표에서 변수와 값 레이블 또는 변수 이름과 값 사용 여부를 지정할 수 있습니다. 레이블의 최대 길이를 지정할 수도 있습니다.

범주형 회귀분석 정규화

방법. 정규화 방법을 사용하면 추정값을 0으로 축소시켜 회귀계수 추정값의 변동을 줄임으로써 모형의 예측 오류를 개선할 수 있습니다. Lasso 및 Elastic Net은 일부 계수 추정값을 정확히 0까지 축소시키고 따라서 변수 형태의 선택을 제공합니다. 정규화 방법이 필요한 경우 각 페널티 계수값에 대해 정규화된 모형과 계수는 현재 세션의 외부 IBM SPSS Statistics 데이터 파일이나 데이터 세트에 기록됩니다. 자세한 정보는 17 페이지의 『범주형 회귀분석: 저장』 주제를 참조하십시오.

- **능형 회귀분석.** 능형 회귀분석은 페널티항이 제공한 계수 곱하기 페널티 계수와 같아지게 하여 계수를 축소할 수 있습니다. 이 계수의 범위는 0(페널티 없음)에서 1까지이며 범위와 증분을 지정할 경우 프로시저에서 "최상"의 페널티 값을 검색합니다.
- **Lasso.** Lasso의 페널티항은 절대 계수의 합을 기준으로 하며 페널티 계수의 지정 사항은 능형 회귀분석과 유사하지만 Lasso가 계산적으로 더 정확합니다.
- **Elastic net.** Elastic Net은 Lasso와 능형 회귀분석 페널티를 간단하게 조합하며 지정된 변수값의 눈금을 검색하여 "최상"의 Lasso 및 Ridge 회귀분석 페널티 계수를 찾습니다. 주어진 Lasso 및 Ridge 회귀분석 페널티의 경우 Elastic Net이 Lasso보다 계산 비용이 더 많이 듭니다.

정규화 도표를 표시합니다. 정규화 도표는 회귀계수 대비 정규화 페널티에 대한 도표입니다. "최상"의 페널티 계수에 대한 범위값을 검색할 경우 해당 범위에서 회귀계수의 변경 방법을 볼 수 있습니다.

Elastic Net 도표. Elastic Net 방법의 경우 능형 회귀분석 벌점값을 기준으로 별도의 정규화 도표가 생성됩니다. 모든 도표는 지정된 최소 및 최대 능형 회귀분석 페널티별로 결정된 범위 내의 모든 값을 사용합니다. 일부 능형 페널티의 경우 최소 및 최대값별로 결정된 값 서브세트를 지정할 수 있습니다. 페널티값 수를 입력하거나 값의 범위를 지정하고 추가를 클릭하면 됩니다.

범주형 회귀분석: 결과

결과 대화 상자를 사용하여 결과에 표시할 통계를 선택할 수 있습니다.

테이블. 다음에 대한 테이블을 작성합니다.

- 다중 R , R^2 , 수정된 R^2 , 최적화 척도법을 고려하는 수정된 R^2 를 포함합니다.
- ANOVA. 이 옵션은 회귀분석 및 잔차 제곱합, 평균 제곱, F 를 포함합니다. 수와 동일한 회귀분석에 대한 자유도를 가지는 분산 분석표와 최적화 척도법을 고려하는 회귀분석에 대한 자유도를 가지는 두 개의 분산 분석표가 표시됩니다.
- 계수. 이 옵션에서 제공하는 테이블은 베타, 베타의 표준 오차, t 값, 유의수준을 포함하는 계수표, 최적화 척도법 자유도를 고려하는 베타의 표준 오차를 갖는 계수 최적화 척도법 표, 0차 상관, 부분상관 및 편상관, 변환된 예측변수에 대한 Pratt의 관계 중요도 측도 및 변환 전후의 허용 오차를 갖는 테이블입니다.
- 반복 히스토리. 알고리즘에 대한 시작값을 포함하여 각 반복에 대해 다중 R 및 회귀분석 오류가 표시됩니다. 첫 번째 반복에서 시작하여 다중 R 의 증가량이 표시됩니다.
- 원 변수간의 상관. 변환되지 않은 변수의 상관을 나타내는 행렬이 표시됩니다.
- 변환된 변수간의 상관. 변환된 변수의 상관을 나타내는 행렬이 표시됩니다.
- 정규화 모형 및 계수 각 정규화 모형의 페널티값, R 제곱 및 회귀계수를 표시합니다. 표본 재추출 방법 또는 보조 오브젝트(검정 케이스)가 지정되어 있는 경우 예측 오차나 검정 MSE도 표시합니다.

표본 재추출. 표본 재추출 방법은 모형에 대한 예측 오차 추정값을 제공합니다.

- 교차 검증. 교차 검증은 표본을 다수의 부표본 또는 fold로 나눕니다. 그런 다음 범주형 회귀 모형을 생성하고 각 부표본에서 차례대로 데이터를 제외합니다. 첫 번째 모형은 첫 번째 표본 중첩의 케이스를 제외한 모든 케이스를 기준으로 하며, 두 번째 모형은 두 번째 표본 중첩의 케이스를 제외한 모든 케이스를 기준으로 하는 방식입니다. 각 모형의 경우 모형 생성에서 제외된 부표본에 적용하여 예측 오차를 추정합니다.
- .632 부트스트랩. 부트스트랩을 사용하여 데이터에서 복원으로 관측값이 무작위로 추출됩니다. 이 프로세스를 몇 번 반복하여 부트스트랩 표본 수를 얻습니다. 모형은 각 부트스트랩 표본에 적합합니다. 각 모형의 예측 오차는 부트스트랩 표본이 아니라 케이스에 맞춰진 모형을 적용하여 추정합니다.

범주 수량화. 선택한 변수의 변환된 값을 보여 주는 테이블이 표시됩니다.

기술 통계. 선택한 변수의 빈도, 결측값, 최빈값을 보여 주는 테이블이 표시됩니다.

범주형 회귀분석: 저장

저장 대화 상자를 사용하면 예측값, 잔차 및 변환된 값을 활성 데이터 세트에 저장할 수 있고, 이산화 데이터, 변환된 값, 정규화된 모형 및 계수, 회귀계수 부호 등을 외부 IBM SPSS Statistics 데이터 파일이나 현재 세션의 데이터 세트에 저장할 수 있습니다.

- 데이터 세트는 현재 세션을 진행하는 동안 사용할 수 있지만 이를 명시적인 데이터 파일로 저장하지 않으면 이후의 세션에서는 사용할 수 없습니다. 데이터 세트 이름은 변수이름 규칙을 준수해야 합니다.
- 파일 이름이나 데이터 세트 이름은 저장된 각 데이터 유형에 따라 서로 달라야 합니다.

정규화 방법 및 계수는 정규화 대화 상자에서 정규화 방법을 선택할 때마다 저장됩니다. 기본적으로 이 프로시저에서는 고유한 이름을 가진 새 데이터 세트를 만들지만 자체 이름을 지정하거나 외부 파일에 기록할 수도 있습니다.

회귀계수 부호는 옵션 대화 상자에서 다중 계통 시작을 초기 구성으로 사용할 때마다 저장됩니다. 기본적으로 이 프로시저에서는 고유한 이름을 가진 새 데이터 세트를 만들지만 자체 이름을 지정하거나 외부 파일에 기록할 수도 있습니다.

범주형 회귀분석 변환 도표

도표 대화 상자를 사용하면 변환 도표와 잔차 도표를 작성하기 위한 변수를 지정할 수 있습니다.

변환 도표. 이러한 변수 각각에 대해 원래 범주 값에 대한 범주 수량화가 도표로 작성됩니다. 빈 범주는 수평축에 표시되지만 계산에는 적용되지 않습니다. 이러한 범주는 수량화에 연결되는 선 구분에 의해 식별됩니다.

잔차 도표. 변수별로 잔차 범주 지표에 대해 해당 예측변수를 제외한 모든 예측변수에서 예측된 종속변수에 대해 계산된 잔차가 도표로 작성되고 범주 표시기에 대해 베타가 곱해진 최적 범주 수량에 대한 잔차가 도표로 작성됩니다.

CATREG 명령 추가 기능

선택을 구문 창에 붙여넣고 CATREG 명령 구문을 편집하여 범주형 회귀분석을 사용자 정의할 수 있습니다. 명령 구문을 사용하여 수행할 수 있는 추가 기능은 다음과 같습니다.

- SAVE 하위 명령을 사용하여 변환된 변수의 루트 이름을 지정하고 이를 활성 데이터 세트에 저장할 수 있습니다.

명령 구문에 대한 자세한 내용은 *Command Syntax Reference*를 참조하십시오.

제 3 장 범주형 주성분 분석(CATPCA)

이 프로시저에서는 데이터의 차원을 축소하는 동시에 범주형 변수를 수량화합니다. 범주형 주성분 분석은 약자로 CATPCA(CATegorical Principal Components Analysis)라고도 합니다.

주성분분석의 목적은 원래 변수에 있는 정보의 대부분을 표현하면서도 원래 변수 세트를 더 작고 상관성이 없는 성분의 세트로 축소하는 데 있습니다. 이 기법은 변수가 많아서 오브젝트(주제와 단위) 사이의 관계를 효과적으로 해석할 수 없을 때 특히 유용합니다. 차원을 축소하여 많은 수의 변수를 해석하는 대신 몇 개의 성분을 해석합니다.

표준 주성분분석에서는 숫자변수 사이의 선형 관계를 가정합니다. 한편 최적화 척도법을 사용하여 변수를 서로 다른 수준으로 척도화할 수 있습니다. 범주형 변수는 지정된 차원에서 최적으로 수량화됩니다. 결과적으로 변수 사이의 비선형 관계도 모델링할 수 있습니다.

예제. 범주형 주성분 분석을 사용하여 직업 범주, 직업 분류, 지역, 여행 횟수(많음, 중간, 적음) 및 직업 만족도 사이의 관계를 설명할 수 있습니다. 두 개의 차원으로 많은 양의 분산을 설명한다는 것을 알 수 있을 것입니다. 첫 번째 차원은 영역에서 직업 범주를 분리하고, 두 번째 차원은 여행 횟수에서 직업 분류를 분리할 수 있습니다. 또한 직업에 대한 높은 만족도는 중간 정도의 여행 횟수와 관련이 있다는 것을 알 수 있습니다.

통계 및 도표. 빈도, 결측값, 최적척도 수준, 최빈값, 중심값 좌표로 설명되는 분산, 벡터 좌표, 변수와 차원에 대한 합계, 벡터 양적변수의 성분 로딩, 범주 수량화와 좌표, 반복 히스토리, 변환된 변수의 상관과 상관 행렬의 고유 값, 원래 변수의 상관과 상관 행렬의 고유 값, 오브젝트 스코어, 범주도표, 결합범주도표, 변환 도표, 잔차 도표, 투영된 중심값 도표, 대상도표, Bi-플롯, Tri-플롯 및 성분 로딩 도표를 선택할 수 있습니다.

범주형 주성분 분석 데이터 고려 사항

데이터. 문자열 변수 값을 항상 문숫자 순서의 오름차순으로 정렬하여 양의 정수로 변환합니다. 사용자 정의 결측값, 시스템 결측값 및 1보다 작은 값을 결측값으로 간주합니다. 비결측으로 처리하려면 코딩을 변경하거나 1 미만의 값을 갖는 변수에 상수를 더할 수 있습니다.

가정. 데이터에는 세 개 이상의 유효 케이스가 있어야 합니다. 분석은 양의 정수 데이터를 기준으로 합니다. 이산화 옵션은 분수 값을 갖는 변수를 "정규"분포에 가까운 분포를 갖는 값으로 그룹화하여 자동으로 범주화합니다. 그리고 문자열 변수의 값을 자동으로 양의 정수로 변환합니다. 다른 이산화 방법을 지정할 수 있습니다.

관련 프로시저. 표준 주성분분석에 해당하는 숫자 수준으로 모든 변수를 척도화합니다. 표준 선형 주성분분석에서 변환된 변수를 사용하여 대체 도표 기능을 사용할 수 있습니다. 모든 변수에 다중 명목 척도 수준이 있는 경우 범주형 주성분 분석은 다중 대응일치 분석과 동일합니다. 변수 세트가 관심 대상인 경우 범주형(비선형) 정준상관분석을 사용해야 합니다.

범주형 주성분 분석 수행 방법

1. 메뉴에서 다음을 선택합니다.

분석 > 차원 축소 > 최적화 척도법...

2. 일부 변수는 다중 명목 아님을 선택합니다.

3. 단일 세트를 선택합니다.

4. 정의를 클릭합니다.

5. 두 개 이상의 분석변수를 선택하고 해법 차원에서 수를 지정합니다.

6. 확인을 클릭합니다.

선택적으로 보조변수를 지정하고 보조변수를 구한 해 또는 도표에 대한 변수 레이블로 적합시킵니다.

CATPCA에서 척도 및 가중치 정의

분석변수와 보조변수의 최적척도 수준을 설정할 수 있습니다. 기본적으로 두 개의 내부 마디가 있는 2차 단조 스플라인(순서)으로 척도화됩니다. 추가적으로 분석변수의 가중치를 설정할 수 있습니다.

변수 가중치. 각 변수에 대한 가중치 정의를 선택할 수 있습니다. 지정한 값은 반드시 양의 정수이어야 합니다. 기본값은 1입니다.

최적척도 수준. 각 변수를 수량화하는 데 사용하는 척도 수준을 선택할 수도 있습니다.

- **스플라인 순서.** 관측 변수에 대한 범주 순서는 최적척도화 변수에 보관됩니다. 범주점은 원점을 지나는 직선(벡터)이 됩니다. 변환 결과는 선택한 차수의 평활 단조 조각별 다항입니다. 조각은 내부 마디에 대한 사용자 정의 수와 프로시저 지정 배치에 따라 지정됩니다.
- **스플라인 명목.** 관측범주의 범주 오브젝트 그룹만 유지하면서 변수를 최적수준으로 척도화합니다. 관측변수의 범주 순서는 유지되지 않습니다. 범주점은 원점을 지나는 직선(벡터)이 됩니다. 변환 결과는 평활이며 단조가 아닐 수도 있는 선택한 차수의 조각별 다항입니다. 조각은 내부 마디에 대한 사용자 정의 수와 프로시저 지정 배치에 따라 지정됩니다.
- **다중 명목.** 관측범주의 범주 오브젝트 그룹만 유지하면서 변수를 최적수준으로 척도화합니다. 관측변수의 범주 순서는 유지되지 않습니다. 범주 점은 특정 범주에서 오브젝트 중심값에 있을 것입니다. 다중은 수량화의 여러 세트를 각 차원에 대하여 구했음을 나타냅니다.
- **순서.** 관측 변수에 대한 범주 순서는 최적척도화 변수에 보관됩니다. 범주점은 원점을 지나는 직선(벡터)이 됩니다. 변환 결과는 스플라인 순서 변환보다 더 잘 적합되지만 평활 정도가 약합니다.
- **명목.** 관측범주의 범주 오브젝트 그룹만 유지하면서 변수를 최적수준으로 척도화합니다. 관측변수의 범주 순서는 유지되지 않습니다. 범주점은 원점을 지나는 직선(벡터)이 됩니다. 변환 결과는 스플라인 명목 변환보다 더 잘 적합되지만 평활의 정도가 약합니다.
- **숫자.** 범주는 동일한 간격(구간 수준)으로 순서화된 것으로 간주됩니다. 범주 순서를 유지하고 관측변수 간 범주를 동일한 간격으로 유지하면서 최적수준으로 척도화합니다. 범주점은 원점을 지나는 직선(벡터)이 됩니다. 모든 변수가 숫자 수준일 경우 분석은 표준 주성분 분석과 유사합니다.

범주형 주성분 분석: 이산화

이산화 대화 상자에서 변수의 코딩을 변경하는 방법을 선택할 수 있습니다. 분수 값 변수는 7개의 범주(또는 이 숫자가 7 미만인 경우 개별값의 수로) 그룹화되는데 이 범주는 특별히 지정하지 않는 한 근사적으로 정규 분포를 따릅니다. 문자열 변수를 문숫자 순서의 오름차순에 따라 범주 표시기를 할당하여 양의 정수로 변환합니다. 문자열 변수의 이산화는 이러한 정수에 적용됩니다. 기본적으로 다른 변수는 변하지 않습니다. 그 다음 이산화 변수를 분석에 사용합니다.

방법. 그룹, 순위, 곱하기 중에서 선택합니다.

- **그룹.** 지정한 범주의 수로 코딩을 변경하거나 구간별로 코딩을 변경합니다.
- **순위.** 케이스의 순위를 정하여 변수를 이산화합니다.
- **곱하기.** 변수의 현재 값을 표준화하고 10을 곱하고, 반올림하고, 가장 작은 이산화 값이 1이 되도록 상수 값을 더합니다.

그룹. 그룹으로 변수를 이산화할 경우 다음 옵션을 사용할 수 있습니다.

- **범주의 수.** 범주의 수를 지정하고 변수값이 이러한 범주에 따라서 근사적으로 정규 분포 또는 균일분포를 따라야 하는지 지정합니다.
- **동일 구간.** 변수를 동일한 크기의 구간으로 정의된 범주로 코딩변경합니다. 구간의 길이를 지정해야 합니다.

범주형 주성분 분석 결측값

결측값 대화 상자에서는 분석변수와 보조변수에 있는 결측값 처리에 대한 처리 방법을 선택할 수 있습니다.

처리 방법. 결측값 제외(수동 처리), 결측값 대치(능동 처리) 또는 해당 변수에 결측값이 있는 오브젝트 제외(목록별 결측값 제외)를 선택합니다.

- **결측값 제외.** 수량화 후에 상관 대치 선택한 변수에서 결측값을 가지는 오브젝트는 이 변수의 분석에서는 사용하지 않습니다. 모든 변수에 수동 처리를 제공하면 모든 변수에 결측값을 가지는 오브젝트는 보조로 처리됩니다. 결과 대화 상자에서 상관을 지정하는 경우에는 분석 후에 원래 변수의 상관에 대해서 변수의 가장 빈도가 높은 범주 또는 최빈값을 결측값에 대치합니다. 최적으로 척도화된 변수의 상관에서 대치법을 선택할 수 있습니다. 최적으로 척도화된 변수의 최빈값으로 결측값을 대체하려면 **최빈값**을 선택합니다. 임시범주를 수량화하여 결측값을 대체하려면 **임시범주**를 선택합니다. 이 변수에서 결측값이 있는 오브젝트는 동일한(임시) 범주에 속하는 것으로 간주됩니다.
- **결측값 대치** 선택한 변수에서 결측값을 가진 오브젝트에는 대치된 값이 있습니다. 대치법을 선택할 수 있습니다. 결측값을 가장 빈도가 높은 범주로 대체하려면 **최빈값**을 선택합니다. 여러 개의 최빈값이 있는 경우에는 가장 작은 범주 표시기가 있는 값을 사용합니다. 결측값을 임시범주의 동일한 수량화로 대체하려면 **임시범주**를 선택합니다. 이 변수에서 결측값이 있는 오브젝트는 동일한(임시) 범주에 속하는 것으로 간주됩니다.
- **해당 변수에 결측값이 있는 오브젝트 제외** 선택한 변수에 결측값이 있는 오브젝트는 분석에서 제외됩니다. 이 처리 방법은 보조변수에서는 사용할 수 없습니다.

범주형 주성분 분석: 옵션

옵션 대화 상자에서 초기 설정 선택, 반복과 수렴 기준 지정, 정규화 방법 선택, 도표 레이블 방법 선택, 보조 오브젝트 지정을 할 수 있습니다.

보조 오브젝트. 보조 오브젝트로 만들 오브젝트의 케이스 수나 오브젝트 범위의 처음과 마지막 케이스 수를 지정한 다음 추가를 클릭합니다. 보조 오브젝트를 모두 지정할 때까지 계속합니다. 오브젝트를 보조로 지정하는 경우에는 해당 오브젝트에 대한 케이스 가중치는 무시됩니다.

정규화 방법. 오브젝트 스코어와 변수를 정규화하는 다섯 가지의 옵션 중 하나를 지정할 수 있습니다. 주어진 분석에서 한 가지 정규화 방법만 사용할 수 있습니다.

- **변수 주성분.** 이 옵션은 변수 간의 연관을 최적화합니다. 오브젝트 공간의 변수 좌표는 성분 로딩입니다(차원이나 오브젝트 스코어 같은 주성분과의 상관). 변수 간의 상관이 주된 관심사일 경우에 유용합니다.
- **오브젝트 주성분.** 이 옵션은 오브젝트 간의 거리를 최적화합니다. 오브젝트 간의 차이나 유사성이 주된 관심사일 경우에 유용합니다.
- **대칭적.** 오브젝트와 변수 사이의 관계가 주된 관심사일 경우에 이 정규화 옵션을 사용합니다.
- **독립적.** 오브젝트 간의 거리와 변수 간의 상관을 분리하여 조사하려면 이 정규화 옵션을 사용합니다.
- **사용자 정의.** $[-1, 1]$ 사이의 닫힌 구간에 속하는 임의의 실수를 지정할 수 있습니다. 1은 오브젝트 주성분 방법, 0은 대칭 방법, -1은 분산 주성분 방법에 해당합니다. -1보다 크고 1보다 작은 값을 지정하여 오브젝트와 변수 둘 다에 대한 고유 값을 전개할 수 있습니다. 이 방법은 특별한 목적의 Bi-플롯 또는 Tri-플롯 작성에 유용합니다.

기준. 프로시저가 계산을 계속할 수 있도록 최대반복수를 지정할 수 있습니다. 또한 수렴 기준 값을 선택할 수 있습니다. 알고리즘은 마지막 두 반복의 전체 적합 차이가 수렴 값보다 작거나 최대반복수에 도달하면 반복을 중단합니다.

도표 레이블기준변수. 도표에서 변수와 값 레이블 또는 변수 이름과 값 사용 여부를 지정할 수 있습니다. 레이블의 최대 길이를 지정할 수도 있습니다.

도표 차원. 결과에 표시되는 차원을 제어할 수 있습니다.

- **해법에 모든 차원 표시.** 해법의 모든 차원이 산점도 행렬에 표시됩니다.
- **차원 수 제한.** 표시되는 차원이 도표 대응으로 제한됩니다. 차원을 제한하는 경우 도표화할 최하위 차원과 최상위 차원을 선택해야 합니다. 최하위 차원의 범위는 1부터 시작하여 해법의 차원 수에서 1을 뺀 값까지 가능하며 더 높은 차원에 대해 도표화됩니다. 최상위 차원 값의 범위는 2부터 시작하여 해법의 차원 수까지 가능하며 차원 쌍을 도표화하는 데 사용할 가장 높은 차원을 가리킵니다. 이 지정 사항은 요청된 모든 다차원 도표에 적용됩니다.

설정. 설정 좌표를 포함하는 파일의 데이터를 읽을 수 있습니다. 파일에 있는 첫 번째 변수에는 첫 번째 차원의 좌표가 있어야 하고, 두 번째 변수에는 두 번째 차원의 좌표가 있어야 하고, 나머지도 마찬가지입니다.

- **초기.** 파일에서 지정한 설정은 분석의 시작점으로 사용됩니다.

- 고정. 파일에서 지정한 설정은 변수 적합으로 사용됩니다. 적합되는 변수는 분석변수로 선택되어야 하지만 설정이 고정되어 있기 때문에 보조변수로 취급되므로 보조변수로 선택할 필요가 없습니다.

범주형 주성분 분석 결과

결과 대화 상자에서 오브젝트 스코어, 성분 로딩, 반복 히스토리, 원 변수 및 변환된 변수간의 상관, 변수와 차원마다 설명된 분산, 선택한 변수에 대한 범주 수량화, 선택된 변수에 대한 기술통계 등을 테이블로 작성할 수 있습니다.

오브젝트 스코어. 오브젝트 스코어를 표시하며 오브젝트 스코어에는 다음과 같은 옵션이 있습니다.

- 다음 범주 포함. 선택한 분석변수의 범주 표시기를 나타냅니다.
- 오브젝트 스코어 레이블 기준. 레이블변수로 지정된 변수 목록에서 오브젝트를 설명할 변수를 하나 선택할 수 있습니다.

성분 로딩. 다중 명목 척도 수준이 없는 모든 변수에 대해 성분 로딩을 나타냅니다.

반복 히스토리. 각 반복에 대하여 설명된 분산, 손실 함수, 설명된 분산의 증가량을 표시합니다.

원 변수간의 상관. 원래 변수의 상관 행렬과 그 행렬의 고유 값을 나타냅니다.

변환된 변수간의 상관. 변환된(최적으로 척도화된) 변수의 상관 행렬과 그 행렬의 고유 값을 나타냅니다.

설명된 분산. 각 변수와 차원마다 중심값 좌표, 벡터 좌표, 중심값 좌표와 벡터 좌표의 결합에 의해서 설명된 분산의 양을 나타냅니다.

범주 수량화. 범주를 수량화하고 선택한 변수의 각 차원에 대한 좌표를 설정합니다.

기술통계. 선택한 변수의 빈도, 결측값의 수, 최빈값을 표시합니다.

범주형 주성분 분석: 저장

저장 대화 상자를 사용하면 이산화 데이터, 오브젝트 스코어, 변환된 값 및 근사값을 외부 IBM SPSS Statistics 데이터 파일 또는 현재 세션의 데이터 세트에 저장할 수 있습니다. 변환된 값, 오브젝트 스코어 및 근사값을 활성 데이터 세트에 저장할 수도 있습니다.

- 데이터 세트는 현재 세션을 진행하는 동안 사용할 수 있지만 이를 명시적인 데이터 파일로 저장하지 않으면 이후의 세션에서는 사용할 수 없습니다. 데이터 세트 이름은 변수이름 규칙을 준수해야 합니다.
- 파일 이름이나 데이터 세트 이름은 저장된 각 데이터 유형에 따라 서로 달라야 합니다.
- 오브젝트 스코어나 변환된 값을 활성 데이터 세트에 저장하는 경우 다중 명목 차원의 수를 지정할 수 있습니다.

범주형 주성분 분석 대상도표

오브젝트 및 변수 도표 대화 상자에서 원하는 도표의 유형과 도표를 작성할 변수를 지정할 수 있습니다.

오브젝트 포인트. 오브젝트 포인트 도표가 나타납니다.

오브젝트 및 변수(Bi-플롯). 오브젝트 포인트는 선택한 변수 좌표(성분 로딩 또는 변수 중심값)에 도표화됩니다.

오브젝트, 로딩 및 중심값(Tri-플롯). 오브젝트 포인트는 다중 명목 척도 수준 변수의 중심값과 다른 변수의 성분 로딩과 함께 도표화됩니다.

Bi-플롯 및 Tri-플롯 변수. Bi-플롯과 Tri-플롯에 대해 모든 변수를 사용하도록 선택하거나 서브세트를 선택할 수 있습니다.

레이블 오브젝트. 오브젝트를 선택한 변수의 범주로 설명하거나(옵션 대화 상자에서 범주 표시기 값 또는 값 레이블을 선택) 케이스의 수로 설명하도록 선택할 수 있습니다. 변수를 선택하면 변수마다 도표 하나가 작성됩니다.

범주형 주성분 분석 범주도표

범주도표 대화 상자에서 원하는 도표의 유형과 도표를 작성할 변수를 지정할 수 있습니다.

범주도표. 선택된 각 변수에 대해서 중심값 도표와 벡터 좌표를 도표화합니다. 다중 명목 척도 수준을 가진 변수에서는 범주가 특정 범주에 있는 오브젝트의 중심값에 있습니다. 다른 모든 척도 수준에서는 범주가 원점을 지나는 벡터 위에 있습니다.

결합범주도표. 선택한 각 변수의 단일 중심값 도표와 벡터 좌표입니다.

변환 도표. 최적 범주 수량화와 범주 표시기의 도표를 나타냅니다. 다중 명목 척도 수준 변수의 차원 수를 지정할 수 있습니다. 각 차원에 대해 하나의 도표를 생성합니다. 선택된 각 변수에 대해 잔차 도표 표시를 선택할 수도 있습니다.

투영할 중심값. 변수를 선택하고 그 중심값을 선택한 변수 위로 투영할 수 있습니다. 다중 명목 척도 수준 변수를 투영할 변수로 선택할 수 없습니다. 이 도표를 요청하는 경우 투영된 중심값 좌표가 있는 테이블이 나타납니다.

범주형 주성분 분석 로딩 도표

로딩 도표 대화 상자에서 도표에 포함할 변수를 지정하고 도표에 중심값을 포함할 것인지 여부를 지정할 수 있습니다.

성분 로딩 표시. 선택한 경우 성분 로딩 도표를 표시합니다.

로딩 변수. 성분 로딩 도표에 대해 모든 변수를 사용하도록 선택하거나 서브세트를 선택할 수 있습니다.

중심값 포함. 다중 명목 척도 수준 변수에는 성분 로딩이 없지만 도표에서는 이러한 변수의 중심값을 포함하도록 선택할 수 있습니다. 다중 명목변수를 모두 사용하도록 선택하거나 서브세트를 선택할 수 있습니다.

CATPCA 명령 추가 기능

선택을 구문 창에 붙여넣고 그에 따른 CATPCA 명령 구문을 편집하는 경우에는 범주형 주성분 분석을 사용자 정의할 수 있습니다. 명령 구문을 사용하여 수행할 수 있는 추가 기능은 다음과 같습니다.

- SAVE 하위 명령을 사용하여 변환된 변수, 오브젝트 스코어 및 근사값의 루트 이름을 지정하고 이를 활성 데이터 세트에 저장할 수 있습니다.
- PLOT 하위 명령을 사용하여 각 도표에 대한 레이블의 최대 길이를 개별적으로 지정할 수 있습니다.
- PLOT 하위 명령을 사용하여 잔차 도표에 대한 개별 변수 목록을 지정할 수 있습니다.

명령 구문에 대한 자세한 내용은 *Command Syntax Reference*를 참조하십시오.

제 4 장 비선형 정준 상관 분석(OVERALS)

비선형 정준 상관 분석은 최적화 척도법을 사용한 범주형 정준상관분석에 해당합니다. 이 프로시저의 목적은 범주형 변수 세트가 서로 간에 어느 정도 유사한지를 확인하는 것입니다. 비선형 정준 상관 분석은 약자로 OVERALS라고도 합니다.

표준 정준상관분석은 두 번째 세트가 단일 반응변수가 아닌 여러 반응변수를 포함하는 다중 회귀분석의 확장입니다. 비선형 정준 상관 분석의 목적은 저차원 공간에 있는 두 개의 연속변수 세트 간의 관계에서 분산 정도를 설명하는 것입니다. 각 세트의 변수는 처음에 선형으로 결합되어 선형 조합이 최대의 상관성을 가지고 있습니다. 이들 조합에서 다음에 이어지는 선형 조합은 이전의 조합과 상관성은 없지만 최대의 상관 가능성은 있음을 알 수 있습니다.

최적화 척도법은 세 가지 중요한 방식으로 표준 분석을 확장합니다. 첫째, OVERALS에서 둘 이상의 변수 세트를 사용할 수 있습니다. 둘째, 변수를 명목, 순서 또는 연속으로 척도화할 수 있습니다. 이로 인해 변수 간의 비선형 관계를 분석할 수 있게 합니다. 마지막으로, 변수 세트 간의 상관성을 최대화하는 대신 변수 세트들을 오브젝트 스코어에 따라 정의되는 알 수 없는 상관 세트와 비교합니다.

예제. 최적화 척도법과 함께 범주형 정준상관분석을 사용하여 직업 범주와 교육 기간을 포함하는 한 변수 세트와 거주 지역과 성별을 포함하는 다른 변수 세트의 관계를 그림으로 나타낼 수 있습니다. 다른 변수보다 교육 기간과 거주 지역 변수가 쉽게 판별되는 것을 알 수 있습니다. 또한 맨 처음 차원에서는 교육 기간 변수가 가장 잘 판별되는 것을 볼 수 있습니다.

통계 및 도표. 빈도, 중심값, 반복 히스토리, 오브젝트 스코어, 범주 수량화, 기중치, 성분 로딩, 단일적합 및 다중 적합값, 오브젝트 스코어도표, 범주 축 도표, 성분 로딩 도표, 범주 중심값 도표, 변환 도표를 구합니다.

비선형 정준 상관 분석 데이터 고려 사항

데이터. 범주형 변수(명목 또는 순서 척도 수준)를 코드화하려면 정수를 사용합니다. 결과를 최소화하려면 1에서 시작하는 연속적인 정수를 사용하여 각 변수를 코딩합니다. 연속 수준으로 척도화된 변수는 연속적인 정수로 코딩변경할 수 없습니다. 결과를 최소화하려면 연속 수준으로 척도화된 각 변수에 대해 모든 값에서 가장 작은 관측값을 빼고 1을 더합니다. 분수에서 소수점 이하 자릿수는 잘립니다.

가정. 변수를 둘 이상의 변수 세트로 분류할 수 있습니다. 분석 과정에서 다중 명목, 단일 명목, 순서 또는 연속으로 변수를 척도화합니다. 이 프로시저에 사용되는 최대 차원 수는 변수의 최적척도 수준에 따라 다릅니다. 모든 변수를 순서, 단일 명목 또는 연속으로 지정한 경우 최대 차원 수는 관측값에서 1을 뺀 값이나 변수의 총계 중에서 작은 것입니다. 그러나 두 개의 변수 세트만 지정한 경우에는 더 작은 세트의 변수 수가 최대차원수입니다. 일부 변수가 다중 명목으로 지정된 경우에는 다중 명목 범주 수의 합계에 비다중 명목변수의 수를 더한 다음 다중 명목변수의 개수를 뺀 값이 최대 차원 수가 됩니다. 예를 들어, 분석에 관련된 변수가 다섯 개이고 그 중 하나는 네 개의 범주가 있는 다중 명목인 경우 최대 차원 수는 $(4 + 4 - 1)$, 즉 7입니다. 최대값보다 큰 차원 수를 지정하는 경우 지정된 값을 무시한 채 최대값이 사용됩니다.

관련 프로시저. 각 변수 세트에 포함된 변수가 하나인 경우에 비선형 정준 상관 분석은 최적화 척도법을 사용하는 주성분 분석과 동일합니다. 이들 각 변수가 다중 명목인 경우 이 분석은 다중 대응일치 분석과 같습니다. 두 개의 변수 세트가 관련되어 있고 변수 세트 중 하나에 변수가 하나만 포함되어 있는 경우 이 분석은 최적화 척도법을 사용하는 범주형 회귀분석과 같습니다.

비선형 정준 상관 분석 수행 방법

1. 메뉴에서 다음을 선택합니다.

분석 > 차원 축소 > 최적화 척도법...

2. 모든 변수가 다중 명목 또는 일부 변수는 다중 명목 이념을 선택합니다.

3. 다중 세트를 선택합니다.

4. 정의를 클릭합니다.

5. 둘 이상의 변수 세트를 정의합니다. 첫 번째 변수 세트에 포함시킬 변수를 선택합니다. 다음을 클릭하여 다음 변수 세트로 이동하고 두 번째 변수 세트에 포함시킬 변수를 선택합니다. 변수 세트를 더 추가할 수 있습니다. 이전에 정의한 변수 세트로 되돌아가려면 이전을 클릭합니다.

6. 선택한 각 변수에 대해 값 범위와 측정 척도(최적화 척도법 수준)를 정의합니다.

7. 확인을 클릭합니다.

8. 또는 다음 작업을 수행합니다.

- 오브젝트 스코어도표에서 점 레이블을 나타내려면 하나 이상의 변수를 선택합니다. 각 변수를 사용하여 해당 변수값에 따라 점을 표시한 개별 도표를 작성합니다. 이러한 도표 레이블 변수 각각에 대한 범위를 지정해야 합니다. 이 대화 상자를 사용하는 경우 단일-변수를 분석에 사용하면서 동시에 레이블 변수로 사용할 수는 없습니다. 분석에 사용된 변수로 오브젝트 스코어도표를 설명해야 하는 경우 변환 메뉴에 있는 변수 계산 기능을 사용하여 해당 변수의 사본을 만듭니다. 도표에 레이블을 표시하려면 새 변수를 사용합니다. 또는 명령 구문을 사용할 수 있습니다.
- 해법에 필요한 차원 수를 지정합니다. 일반적으로 대부분의 변동 설명에 필요한 만큼의 차원을 선택합니다. 3차원 이상인 분석의 경우 처음 3차원에 대해 3차원 도표를 작성합니다. 차트를 편집하여 다른 차원을 표시할 수 있습니다.

범위 및 척도 정의

각 변수의 범위를 정의해야 합니다. 최대값은 정수로 지정해야 합니다. 분수 데이터 값은 분석에서 잘립니다. 지정한 범위를 벗어난 범주값은 분석에서 무시됩니다. 결과를 최소화하려면 변환 메뉴에 있는 자동 코딩변경 기능을 사용하여 명목 또는 순서로 처리된 변수에 대해 1부터 시작하는 연속적인 범주를 만듭니다. 연속 수준에서 척도화된 변수의 경우 연속적인 정수로 코딩변경하지 않는 것이 좋습니다. 연속으로 처리된 변수에 대한 결과를 최소화하려면 모든 값에서 최소값을 뺀 다음 1을 더합니다.

각 변수를 수량화하는 데 사용되는 척도도 선택해야 합니다.

- 순서. 관측변수의 범주 순서는 양적변수로 유지됩니다.
- 단일 명목. 양적변수에서 동일 범주의 오브젝트는 동일 스코어를 받습니다.

- 다중 명목. 각 차원에 대한 수량화가 다를 수 있습니다.
- 이산형 숫자. 범주는 순서대로 동일한 간격으로 처리됩니다. 관측된 변수에 대한 범주 수와 범주 순서 간의 차이는 양적변수로 유지됩니다.

범위지정

각 변수의 범위를 정의해야 합니다. 최대값은 정수로 지정해야 합니다. 분수 데이터 값은 분석에서 잘립니다. 지정한 범위를 벗어난 범주값은 분석에서 무시됩니다. 결과를 최소화하려면 변환 메뉴에 있는 자동 코딩변경 기능을 사용하여 1부터 시작하는 연속적인 범주를 만듭니다.

오브젝트 스코어도표를 설명하는 데 사용되는 각 변수의 범위도 지정해야 합니다. 그러나 변수에 대해 정의된 범위를 벗어난 데이터 값으로 범주를 설명하더라도 이 레이블이 도표에 나타납니다.

비선형 정준 상관 분석 옵션

옵션 대화 상자를 사용하면 선택적 통계와 도표를 선택할 수 있고, 오브젝트 스코어를 활성 데이터 세트에 새 변수로 저장할 수 있고, 반복 및 수렴 기준을 지정할 수 있고, 분석을 위한 초기설정을 지정할 수 있습니다.

출력. 사용할 수 있는 통계에는 주변 빈도(개수), 중심값, 반복 히스토리, 가중치 및 성분 로딩, 범주 수량화, 오브젝트 스코어, 단일적합 및 다중 적합값 통계가 포함됩니다.

- 중심값(*Centroids*). 범주 수량화로서 오브젝트(케이스)에 대한 투영된 오브젝트 스코어의 실제 평균은 변수의 동일 범주에 속하는 케이스의 각 세트에 포함됩니다.
- 가중치 및 성분 로딩(*Weights and component loadings*). 오브젝트 스코어가 양적변수로 회귀되는 세트의 모든 양적변수에 대한 각 차원과 오브젝트 공간에 있는 양적변수의 투영에 있는 회귀계수입니다. 각 변수가 각 세트 내의 차원에 미치는 영향력을 나타냅니다.
- 단일적합 및 다중 적합값(*Single and multiple fit*). 단일 및 다중 범주 좌표/오브젝트별 범주 수량화의 적합도에 대한 측도입니다.
- 범주 수량화(*Category quantifications*). 변수 범주에 최적 척도값을 할당합니다.
- 오브젝트 스코어(*Object scores*). 특정 차원의 오브젝트(케이스)에 할당된 최적 스코어입니다.

도표. 범주 좌표, 오브젝트 스코어, 성분 로딩, 범주 중심값 및 변환 도표를 작성할 수 있습니다.

오브젝트 스코어 저장. 오브젝트 스코어를 활성 데이터 세트에 새 변수로 저장할 수 있습니다. 오브젝트 스코어는 기본 대화 상자에서 지정한 차원 수에 대해 저장됩니다.

무작위 초기설정 사용. 변수 전체 또는 일부가 단일 명목인 경우에는 무작위 초기설정을 사용해야 합니다. 이 옵션을 선택하지 않으면 중첩된 초기설정이 사용됩니다.

기준. 계산을 반복하는 비선형 정준 상관 분석의 최대반복수를 지정할 수 있습니다. 또한 수렴 기준 값을 선택할 수 있습니다. 마지막 두 번의 반복 간에 전체 적합 차이가 수렴값 미만이거나 최대반복수에 도달한 경우에는 분석이 중단됩니다.

OVERALS 명령 추가 기능

선택을 구문 창으로 붙여넣고 결과로 표시된 OVERALS 명령 구문을 편집하는 경우 비선형 정준 상관 분석을 사용자 정의할 수 있습니다. 명령 구문을 사용하여 수행할 수 있는 추가 기능은 다음과 같습니다.

- PLOT 하위 명령에서 NDIM 키워드를 사용하여 추출된 모든 차원을 도표화하지 않고 도표화할 차원 쌍을 지정할 수 있습니다.
- PLOT 하위 명령을 사용하여 도표의 점을 설명하는 데 사용되는 값 레이블 문자의 수를 지정할 수 있습니다.
- PLOT 하위 명령을 사용하여 오브젝트 스코어도표의 레이블 변수로 여섯 개 이상의 변수를 지정할 수 있습니다.
- PLOT 하위 명령을 사용하여 오브젝트 스코어도표에 대한 레이블 변수로 분석에 사용되는 변수를 선택할 수 있습니다.
- PLOT 하위 명령을 사용하여 수량화 스코어 도표의 점 레이블을 제공하는 변수를 선택할 수 있습니다.
- 활성 데이터 세트의 케이스를 일부만 사용하려는 경우 NOBSERVATIONS 하위 명령을 사용하여 분석에 포함할 케이스 수를 지정할 수 있습니다.
- SAVE 하위 명령을 사용하여 오브젝트 스코어 저장 시 생성되는 변수의 루트 이름을 지정할 수 있습니다.
- SAVE 하위 명령을 사용하여 추출된 모든 차원을 저장하지 않고 저장할 차원 수를 지정할 수 있습니다.
- MATRIX 하위 명령을 사용하여 범주 수량화를 행렬 파일로 저장할 수 있습니다.
- SET 명령을 사용하여 일반적인 고해상도 도표보다 쉽게 읽을 수 있는 저해상도 도표를 작성할 수 있습니다.
- PLOT 하위 명령을 사용하여 지정된 변수만을 위한 중심값 도표와 변환 도표를 작성할 수 있습니다.

명령 구문에 대한 자세한 내용은 *Command Syntax Reference*를 참조하십시오.

제 5 장 대응일치 분석

대응일치 분석의 목적 중 하나는 저차원 공간에서 대응일치표를 사용하여 두 명목변수 간의 관계를 설명하는 동시에 각 변수의 범주 관계를 설명하는 것입니다. 각 변수에 대해 도표에서 범주점 사이의 거리는 서로 인접되어 있는 동일한 범주의 관계를 반영합니다. 원점에서 다른 변수의 범주점까지 벡터 상에 위치한 변수에 대한 투영점은 변수간의 관계를 설명합니다.

분할표 분석에는 대개 열 프로파일과 행 프로파일 탐색 및 카이제곱 통계를 사용한 독립성 검정이 포함됩니다. 그러나 프로파일의 수가 상당히 많아져서 카이제곱 검정을 통해 종속 구조가 나타나지 않을 수 있습니다. 교차 분석표 프로시저에서는 여러 연관 측도법과 연관 검정법을 제공하지만 변수 간의 모든 연관을 도표로 설명할 수는 없습니다.

요인 분석은 저차원 공간에서 변수 간의 관계를 설명할 수 있는 표준 기법입니다. 그러나 요인 분석에서는 구간 데이터가 필요하며 관측값 수는 변수 수의 5배가 되어야 합니다. 반면에 대응일치 분석은 명목변수를 가정하여 각 변수에 대한 범주 간의 관계 및 변수 간의 관계를 설명할 수 있습니다. 또한 대응일치 분석은 양의 대응일치 측도표 분석에도 사용할 수 있습니다.

예제. 대응일치 분석을 사용하여 직원 범주와 흡연 습관의 관계를 그림으로 나타낼 수 있습니다. 흡연과 관련하여 주임은 비서와 다르지만 비서는 과장과 다르지 않다는 것을 확인할 수 있을 것입니다. 또한 담배를 많이 피우는 것은 주임과 관계가 있는 반면 적게 피우는 것은 비서와 관계가 있다는 것을 확인할 수 있습니다.

통계 및 도표. 대응일치 측도, 열 프로파일과 행 프로파일, 비정칙 값, 열 스코어와 행 스코어, 관성, 매스, 열 스코어와 행 스코어 신뢰통계, 비정칙 값 신뢰도 통계, 변환 도표, 행 점 도표, 열 점 도표 및 Bi-플롯 등이 있습니다.

대응일치 분석 데이터 고려 사항

데이터. 분석되는 범주형 변수는 명목상으로 척도화됩니다. 통합된 데이터 또는 빈도분석 이외의 대응일치 측도의 경우에는 양의 유사성 값이 있는 가중 변수를 사용합니다. 테이블 데이터의 경우 구문을 사용하여 테이블을 읽습니다.

가정. 프로시저에 사용하는 최대 차원 수는 활성 행과 열 범주 수 및 등제한조건 수에 따라 다릅니다. 등제한 조건을 사용하지 않고 모든 범주를 활성화할 경우 최대 차원 수는 가장 적은 범주를 가진 변수 범주보다 한 차원 작습니다. 예를 들어, 한 변수에는 다섯 개의 범주가 있고 나머지 변수에는 네 개의 범주가 있는 경우 최대 차원 수는 3입니다. 보조 범주는 활성화되지 않습니다. 예를 들어, 한 변수에 다섯 개의 범주가 있고 이들 중 두 개는 보조 범주이고 나머지 변수에는 네 개의 범주가 있는 경우 최대 차원 수는 2입니다. 제약되어 있는 모든 범주 세트를 동일하게 하나의 범주로 처리합니다. 예를 들어, 한 변수에는 다섯 개의 범주가 있고 이들 중 세 개의 범주가 동일하게 제약되어 있으면 최대 차원 수 결정시 해당 변수에는 세 개의 범주가 있는 것으로 처리해야 합니다. 두 개의 범주는 제약되어 있지 않으므로 세 번째 범주는 세 개의 제한된 범주와 대응됩니다. 최대값을 초과하는 차원 수를 지정하는 경우라도 최대값을 사용합니다.

관련 프로시저. 셋 이상의 변수가 관련되어 있는 경우에는 다중 대응일치 분석을 사용합니다. 변수를 순서대로 척도화해야 하는 경우에는 범주형 주성분 분석을 사용합니다.

대응일치 분석 수행 방법

1. 메뉴에서 다음을 선택합니다.

분석 > 차원 축소 > 대응일치 분석...

2. 행 변수를 선택합니다.
3. 열 변수를 선택합니다.
4. 변수의 범위를 지정합니다.
5. 확인을 클릭합니다.

대응일치 분석에서 행 범가지정

행 변수의 범위를 지정해야 합니다. 최소값과 최대값은 정수이어야 합니다. 분수 데이터 값은 분석에서 잘립니다. 지정한 범위를 벗어난 범주값은 분석에서 무시됩니다.

모든 범주는 처음에 제한조건이 없는 상태로 활성화됩니다. 행 범주를 다른 행 범주와 동일하게 제약하거나 보조 범주로 지정할 수 있습니다.

- **동일한 범주.** 범주에는 동일한 스코어가 있어야 합니다. 범주의 순서가 부적당하거나 직관적이지 않은 경우 등제한조건을 사용합니다. 동일하게 제한할 수 있는 최대 행 범주 수는 활성 행 범주의 전체 수에서 1을 뺀 것입니다. 범주 세트에 다른 등제한조건을 설정하려면 구문을 사용합니다. 예를 들어, 범주 1과 범주 2를, 범주 3과 범주 4를 각각 동일한 범주로 제한하려면 구문을 사용합니다.
- **보조 범주.** 보조 범주는 분석에는 영향을 주지 않지만 활성 범주로 정의한 공간에 표시됩니다. 보조 범주는 차원을 정의하는데 필요하지 않습니다. 보조 행 범주의 최대 수는 행 범주의 전체 수에서 2를 뺀 수입니다.

대응일치 분석에서 열 범가지정

열 변수의 범위를 지정해야 합니다. 최소값과 최대값은 정수이어야 합니다. 분수 데이터 값은 분석에서 잘립니다. 지정한 범위를 벗어난 범주값은 분석에서 무시됩니다.

모든 범주는 처음에 제한조건이 없는 상태로 활성화됩니다. 열 범주를 다른 열 범주와 동일하게 제약하거나 보조 범주로 지정할 수 있습니다.

- **동일한 범주.** 범주에는 동일한 스코어가 있어야 합니다. 범주의 순서가 부적당하거나 직관적이지 않은 경우 등제한조건을 사용합니다. 동일하게 제한할 수 있는 최대 열 범주 수는 활성 열 범주의 전체 수에서 1을 뺀 것입니다. 범주 세트에 다른 등제한조건을 설정하려면 구문을 사용합니다. 예를 들어, 범주 1과 범주 2를, 범주 3과 범주 4를 각각 동일한 범주로 제한하려면 구문을 사용합니다.
- **보조 범주.** 보조 범주는 분석에는 영향을 주지 않지만 활성 범주로 정의한 공간에 표시됩니다. 보조 범주는 차원을 정의하는데 필요하지 않습니다. 보조 열 범주의 최대 수는 활성 열 범주의 전체 수에서 2를 뺀 수입니다.

대응일치 분석 모형

모형 대화 상자를 사용하여 차원 수, 거리 척도, 표준화 방법, 정규화 방법을 지정할 수 있습니다.

해법 차원. 차원 수를 지정합니다. 일반적으로 대부분의 변동 설명에 필요한 만큼의 차원을 선택합니다. 최대 차원 수는 분석에 사용되는 활성 범주 수와 등제한조건에 따라 다릅니다. 최대 차원 수는

- 활성 행 범주 수에서 등제한된 행 범주 수를 빼 수에 제한된 행 범주 세트 수를 더한 것보다 작고,
- 활성 열 범주 수에서 등제한된 열 범주 세트 수를 빼 수에 제한된 열 수를 더한 것보다 작습니다.

거리 척도. 대응일치표의 행과 열 사이의 거리 척도를 선택할 수 있습니다. 다음 선택 사항 중 하나를 선택합니다.

- **카이제곱.** 가중 프로파일 거리를 사용합니다. 가중치는 열 또는 행의 매스입니다. 이 척도는 표준 대응일치 분석에 필요합니다.
- **유클리디안.** 대응 열과 대응 행 간의 차이 제곱값 합의 제곱근을 사용합니다.

표준화 방법. 다음 선택 사항 중 하나를 선택합니다.

- **행과 열 평균 제거.** 행과 열 모두 가운데에 위치합니다. 이 방법은 표준 대응일치 분석에 필요합니다.
- **행 평균 제거.** 행만 가운데에 위치합니다.
- **열 평균 제거.** 열만 가운데에 위치합니다.
- **행 전체 균등화 및 평균 제거.** 행을 가운데로 배치하기 전에 행 여백을 균등화합니다.
- **열 전체 균등화 및 평균 제거.** 열을 가운데로 배치하기 전에 열 여백을 균등화합니다.

정규화 방법. 다음 선택 사항 중 하나를 선택합니다.

- **대칭적.** 각 차원에서 행 스코어는 일치 비정칙 값으로 나누어지는 열 스코어의 가중 평균이고 열 스코어는 일치 비정칙 값으로 나누어지는 행 스코어의 가중 평균입니다. 이 방법은 두 변수 범주간의 차이와 유사성을 확인하는 경우에만 사용합니다.
- **주성분.** 행 점과 열 점 사이의 거리는 선택한 거리 척도에 따라 대응일치표에 나타나는 거리의 근사값입니다. 이 방법은 두 변수의 차이 대신 하나 또는 두 범주 간의 차이를 확인하는 경우에 사용합니다.
- **행 주성분.** 행 점 사이의 거리는 선택한 거리 척도에 따라 대응일치표에 나타나는 거리의 근사값입니다. 행 스코어는 열 스코어의 가중 평균입니다. 이 방법은 행 변수 범주 간의 차이와 유사성을 확인하는 경우에 사용합니다.
- **열 주성분.** 열 점 간의 거리는 선택한 거리 척도에 따라 대응일치표에 나타나는 거리의 근사값입니다. 열 스코어는 행 스코어의 가중 평균입니다. 이 방법은 열 변수 범주 간의 차이와 유사성을 확인하는 경우에 사용합니다.
- **사용자 정의.** -1에서 1사이의 값을 지정해야 합니다. -1 값은 열 주성분과 대응합니다. 1 값은 행 주성분과 대응합니다. 0 값은 대칭에 대응합니다. 다른 모든 값은 행과 열 스코어 이상으로 여러 각도로 관성을 펼칩니다. 이 방법은 사용자 정의 Bi-플롯 작성에 유용합니다.

대응일치 분석: 통계

통계 대화 상자를 사용하여 작성되는 숫자 결과를 지정할 수 있습니다.

대응일치표. 행과 열 주변 합계가 있는 입력변수의 교차 분석표입니다.

행 점 개요. 각 행 범주에 대한 스코어, 매스, 관성, 차원 관성에 대한 기여도, 점 관성에 대한 차원의 기여도입니다.

열 점 개요. 각 열 범주에 대한 스코어, 매스, 관성, 차원 관성에 대한 기여도, 점 관성에 대한 차원의 기여도입니다.

행 프로파일. 각 행 범주에 대한 열 변수 범주의 분산입니다.

열 프로파일. 각 열 범주에 대한 행 변수 범주의 분산입니다.

대응일치표의 순열. 대응일치표가 재정리되어 행과 열이 첫 번째 차원의 스코어에 따라 순서가 늘어납니다. 선택적으로 순열표가 생성되는 최대 차원 수를 지정할 수 있습니다. 1부터 지정된 숫자까지 각 차원에 대한 순열표가 작성됩니다.

행 점에 대한 신뢰도 통계. 보조 범주가 아닌 모든 행 점에 대한 표준 편차와 상관을 포함합니다.

열 점에 대한 신뢰도 통계. 보조 범주가 아닌 모든 열 점에 대한 표준 편차와 상관을 포함합니다.

대응일치 분석 도표

도표 대화 상자를 사용하여 작성되는 도표를 지정할 수 있습니다.

산점도. 특정 차원의 모든 대응별 도표 행렬을 작성합니다. 사용할 수 있는 산점도는 다음과 같습니다.

- **Bi-플롯.** 행 점과 열 점이 결합된 도표 행렬을 작성합니다. 주성분 정규화를 선택하면 Bi-플롯을 사용할 수 없습니다.
- **행 점.** 행 점 도표 행렬을 작성합니다.
- **열 점.** 열 점 도표 행렬을 작성합니다.

선택적으로 점 레이블에 사용할 값 레이블의 문자수를 지정할 수 있습니다. 이 값은 20 이하의 양의 정수로 지정해야 합니다.

선형 도표. 선택한 변수에 대한 모든 차원의 도표를 작성합니다. 사용할 수 있는 선형 도표는 다음과 같습니다.

- **변환된 행 범주.** 대응 행 스코어와 반대되는 원래 행 범주 값의 도표를 작성합니다.
- **변환된 열 범주.** 대응 열 스코어와 반대되는 원래 열 범주 값의 도표를 작성합니다.

선택적으로 범주 축 레이블시 사용할 값 레이블의 문자수를 지정할 수 있습니다. 이 값은 20 이하의 양의 정수로 지정해야 합니다.

도표 차원. 결과에 표시되는 차원을 제어할 수 있습니다.

- 해법에 모든 차원 표시. 해법의 모든 차원이 산점도 행렬에 표시됩니다.
- 차원 수 제한. 표시되는 차원이 도표 대응으로 제한됩니다. 차원을 제한하는 경우 도표화할 최하위 차원과 최상위 차원을 선택해야 합니다. 최하위 차원의 범위는 1부터 시작하여 해법의 차원 수에서 1을 뺀 값까지 가능하며 더 높은 차원에 대해 도표화됩니다. 최상위 차원 값의 범위는 2부터 시작하여 해법의 차원 수까지 가능하며 차원 쌍을 도표화하는 데 사용할 가장 높은 차원을 가리킵니다. 이 지정 사항은 요청된 모든 다차원 도표에 적용됩니다.

CORRESPONDENCE 명령 추가 기능

선택을 구문 창에 붙여넣고 CORRESPONDENCE 명령 구문을 편집하여 대응일치 분석을 사용자 정의할 수 있습니다. 명령 구문을 사용하여 수행할 수 있는 추가 기능은 다음과 같습니다.

- TABLE = ALL 하위 명령을 사용하여 케이스별 데이터를 사용하지 않고 입력 값으로 테이블 데이터를 지정할 수 있습니다.
- PLOT 하위 명령을 사용하여 각 산점도 행렬 또는 Bi-플롯 행렬 유형에 대한 점 레이블에 사용되는 값 레이블 문자수를 지정할 수 있습니다.
- PLOT 하위 명령을 사용하여 각 선형 도표 유형에 대한 점 레이블에 사용되는 값 레이블 문자수를 지정할 수 있습니다.
- OUTFILE 하위 명령을 사용하여 행렬 데이터 파일에 행 스코어와 열 스코어 행렬을 작성할 수 있습니다.
- OUTFILE 하위 명령을 사용하여 행렬 데이터 파일에 비정칙 값과 스코어에 대한 신뢰통계(분산 및 공분산) 행렬을 작성할 수 있습니다.
- EQUAL 하위 명령을 사용하여 동일한 다중 범주 세트를 지정할 수 있습니다.

명령 구문에 대한 자세한 내용은 *Command Syntax Reference*를 참조하십시오.

제 6 장 다중 대응일치 분석

다중 대응일치 분석에서는 동일한 범주 내의 오브젝트가 근접하여 함께 배치되고 서로 다른 범주의 오브젝트는 멀리 떨어져 배치되도록 케이스(오브젝트)와 범주에 숫자값을 할당하여 명목(범주형) 데이터를 수량화합니다. 각 오브젝트는 오브젝트에 적용되는 범주의 범주 점에 가능한 한 가깝게 배치됩니다. 이와 같은 방식을 통해 범주는 오브젝트를 동종 하위 그룹으로 분류합니다. 변수는 동일한 범주의 오브젝트를 동일한 하위 그룹으로 분류하는 경우 동일군으로 간주됩니다.

예제. 다중 대응일치 분석을 사용하여 직업 범주, 하위 분류 및 성별 사이의 관계를 그래픽으로 표시할 수 있습니다. 사람들 사이에 소수 분류 및 성별은 구분되지만 직업 범주는 구분되지 않는 것을 확인할 수 있습니다. 또한 라틴아메리카계 및 미국 흑인 범주가 동일하게 나타나는 것을 알 수 있습니다.

통계 및 도표. 오브젝트 스코어, 판별 측도, 반복 히스토리, 원래 변수와 변환된 변수의 상관, 범주 수량화, 기술통계, 오브젝트 포인트 도표, Bi-플롯, 범주도표, 결합범주도표, 변환 도표 및 판별 측도 도표 등이 있습니다.

다중 대응일치 분석 데이터 고려 사항

데이터. 문자열 변수 값을 항상 문숫자 순서의 오름차순으로 정렬하여 양의 정수로 변환합니다. 사용자 정의 결측값, 시스템 결측값 및 1보다 작은 값을 결측값으로 간주합니다. 비결측으로 처리하려면 코딩을 변경하거나 1 미만의 값을 갖는 변수에 상수를 더할 수 있습니다.

가정. 모든 변수는 다중 명목 척도 수준을 갖습니다. 데이터에는 세 개 이상의 유효 케이스가 있어야 합니다. 분석은 양의 정수 데이터를 기준으로 합니다. 이산화 옵션은 분수 값을 갖는 변수의 값을 정규 분포에 가까운 분포를 갖는 범주로 그룹화하여 해당 변수를 자동으로 범주화하고 문자열 변수의 값을 자동으로 양의 정수로 변환합니다. 다른 이산화 방법을 지정할 수 있습니다.

관련 프로시저. 변수가 둘인 경우 다중 대응일치 분석은 대응일치 분석과 같습니다. 변수에 순서 또는 명목 특성이 있다면 범주형 주성분 분석을 사용해야 합니다. 변수 세트가 관심의 대상이라면 비선형 정준 상관 분석을 사용해야 합니다.

다중 대응일치 분석 수행 방법

1. 메뉴에서 다음을 선택합니다.

분석 > 차원 축소 > 최적화 척도법...

2. 모든 변수가 다중 명목을 선택합니다.

3. 단일 세트를 선택합니다.

4. 정의를 클릭합니다.

5. 두 개 이상의 분석변수를 선택하고 해법 차원에서 수를 지정합니다.

6. 확인을 클릭합니다.

선택적으로 보조변수를 지정하고 보조변수를 구한 해 또는 도표에 대한 변수 레이블로 적합시킵니다.

다중 대응일치 분석에서 변수 가중치 정의

분석 변수의 가중치를 설정할 수 있습니다.

변수 가중치. 각 변수에 대한 가중치 정의를 선택할 수 있습니다. 지정한 값은 반드시 양의 정수이어야 합니다. 기본값은 1입니다.

다중 대응일치 분석 이산화

이산화 대화 상자에서 변수의 코딩을 변경하는 방법을 선택할 수 있습니다. 분수값 변수는 일곱 개의 범주로 (또는 이 숫자가 7 미만인 경우 개별값의 수로) 그룹화되는데 이 범주는 특별히 지정하지 않는 한 근사적으로 정규 분포를 따릅니다. 문자열 변수를 문숫자 순서의 오름치순에 따라 범주 표시기를 할당하여 양의 정수로 변환합니다. 문자열 변수의 이산화는 이러한 정수에 적용됩니다. 기본적으로 다른 변수는 변하지 않습니다. 그 다음 이산화 변수를 분석에 사용합니다.

방법. 그룹, 순위, 곱하기 중에서 선택합니다.

- **그룹.** 지정한 범주의 수로 코딩을 변경하거나 구간별로 코딩을 변경합니다.
- **순위.** 케이스의 순위를 정하여 변수를 이산화합니다.
- **곱하기.** 변수의 현재 값을 표준화하고 10을 곱하고, 반올림하고, 가장 작은 이산화 값이 1이 되도록 상수 값을 더합니다.

그룹. 그룹으로 변수를 이산화할 경우 다음 옵션을 사용할 수 있습니다.

- **범주의 수.** 범주의 수를 지정하고 변수값이 이러한 범주에 따라서 근사적으로 정규 분포 또는 균일분포를 따라야 하는지 지정합니다.
- **동일 구간.** 변수를 동일한 크기의 구간으로 정의된 범주로 코딩변경합니다. 구간의 길이를 지정해야 합니다.

다중 대응일치 분석 결측값

결측값 대화 상자에서는 분석변수와 보조변수에 있는 결측값 처리에 대한 처리 방법을 선택할 수 있습니다.

결측값 처리 방법. 결측값 제외(수동 처리), 결측값 대치(능동 처리) 또는 해당 변수에 결측값이 있는 오브젝트 제외(목록별 결측값 제외)를 선택합니다.

- **결측값 제외.** 수량화 후에 상관 대치 선택한 변수에서 결측값을 가지는 오브젝트는 이 변수의 분석에서는 사용하지 않습니다. 모든 변수에 수동 처리를 제공하면 모든 변수에 결측값을 가지는 오브젝트는 보조로 처리됩니다. 결과 대화 상자에서 상관을 지정하는 경우에는 분석 후에 원래 변수의 상관에 대해서 변수의 가장 빈도가 높은 범주 또는 최빈값을 결측값에 대치합니다. 최적으로 척도화된 변수의 상관에서 대치법을 선택할 수 있습니다. 최적으로 척도화된 변수의 최빈값으로 결측값을 대체하려면 최빈값을 선택합니다. 임시범주를 수량화하여 결측값을 대체하려면 임시범주를 선택합니다. 이 변수에서 결측값이 있는 오브젝트는 동일한(임시) 범주에 속하는 것으로 간주됩니다.

- **결측값 대체** 선택한 변수에서 결측값을 가진 오브젝트에는 대체된 값이 있습니다. 대체법을 선택할 수 있습니다. 결측값을 가장 빈도가 높은 범주로 대체하려면 **최빈값**을 선택합니다. 여러 개의 최빈값이 있는 경우에는 가장 작은 범주 표시기가 있는 값을 사용합니다. 결측값을 임시범주의 동일한 수량화로 대체하려면 임시 범주를 선택합니다. 이 변수에서 결측값이 있는 오브젝트는 동일한(임시) 범주에 속하는 것으로 간주됩니다.
- **해당 변수에 결측값이 있는 오브젝트 제외** 선택한 변수에 결측값이 있는 오브젝트는 분석에서 제외됩니다. 이 처리 방법은 보조변수에서는 사용할 수 없습니다.

다중 대응일치 분석 옵션

옵션 대화 상자에서 초기 설정 선택, 반복과 수렴 기준 지정, 정규화 방법 선택, 도표 레이블 방법 선택, 보조 오브젝트 지정을 할 수 있습니다.

보조 오브젝트. 보조 오브젝트로 만들 오브젝트의 케이스 수나 오브젝트 범위의 처음과 마지막 케이스 수를 지정한 다음 추가를 클릭합니다. 보조 오브젝트를 모두 지정할 때까지 계속합니다. 오브젝트를 보조로 지정하는 경우에는 해당 오브젝트에 대한 케이스 가중치는 무시됩니다.

정규화 방법. 오브젝트 스코어와 변수를 정규화하는 다섯 가지의 옵션 중 하나를 지정할 수 있습니다. 주어진 분석에서 한 가지 정규화 방법만 사용할 수 있습니다.

- **변수 주성분.** 이 옵션은 변수 간의 연관을 최적화합니다. 오브젝트 공간의 변수 좌표는 성분 로딩입니다(차원이나 오브젝트 스코어 같은 주성분과의 상관). 변수 간의 상관이 주된 관심사일 경우에 유용합니다.
- **오브젝트 주성분.** 이 옵션은 오브젝트 간의 거리를 최적화합니다. 오브젝트 간의 차이나 유사성이 주된 관심사일 경우에 유용합니다.
- **대칭적.** 오브젝트와 변수 사이의 관계가 주된 관심사일 경우에 이 정규화 옵션을 사용합니다.
- **독립적.** 오브젝트 간의 거리와 변수 간의 상관을 분리하여 조사하려면 이 정규화 옵션을 사용합니다.
- **사용자 정의.** $[-1, 1]$ 사이의 닫힌 구간에 속하는 임의의 실수를 지정할 수 있습니다. 1은 오브젝트 주성분 방법, 0은 대칭 방법, -1은 분산 주성분 방법에 해당합니다. -1보다 크고 1보다 작은 값을 지정하여 오브젝트와 변수 둘 다에 대한 고유 값을 전개할 수 있습니다. 이 방법은 특별한 목적의 Bi-플롯 또는 Tri-플롯 작성에 유용합니다.

기준. 프로시저가 계산을 계속할 수 있도록 최대반복수를 지정할 수 있습니다. 또한 수렴 기준 값을 선택할 수 있습니다. 알고리즘은 마지막 두 반복의 전체 적합 차이가 수렴 값보다 작거나 최대반복수에 도달하면 반복을 중단합니다.

도표 레이블기준변수. 도표에서 변수와 값 레이블 또는 변수 이름과 값 사용 여부를 지정할 수 있습니다. 레이블의 최대 길이를 지정할 수도 있습니다.

도표 차원. 결과에 표시되는 차원을 제어할 수 있습니다.

- **해법에 모든 차원 표시.** 해법의 모든 차원이 산점도 행렬에 표시됩니다.
- **차원 수 제한.** 표시되는 차원이 도표 대응으로 제한됩니다. 차원을 제한하는 경우 도표화할 최하위 차원과 최상위 차원을 선택해야 합니다. 최하위 차원의 범위는 1부터 시작하여 해법의 차원 수에서 1을 뺀 값까지

가능하며 더 높은 차원에 대해 도표화됩니다. 최상위 차원 값의 범위는 2부터 시작하여 해법의 차원 수까지 가능하며 차원 쌍을 도표화하는 데 사용할 가장 높은 차원을 가리킵니다. 이 지정 사항은 요청된 모든 다차원 도표에 적용됩니다.

설정. 설정 좌표를 포함하는 파일의 데이터를 읽을 수 있습니다. 파일에 있는 첫 번째 변수에는 첫 번째 차원의 좌표가 있어야 하고, 두 번째 변수에는 두 번째 차원의 좌표가 있어야 하고, 나머지도 마찬가지입니다.

- 초기. 파일에서 지정한 설정은 분석의 시작점으로 사용됩니다.
- 고정. 파일에서 지정한 설정은 변수 적합으로 사용됩니다. 적합되는 변수는 분석변수로 선택되어야 하지만 설정이 고정되어 있기 때문에 보조변수로 취급되므로 보조변수로 선택할 필요가 없습니다.

다중 대응일치 분석 결과

결과 대화 상자에서 오브젝트 스코어, 판별 측도, 반복 히스토리, 원래 변수와 변환된 변수 간의 상관, 선택된 변수에 대한 범주 수량화, 선택된 변수에 대한 기술통계 등을 테이블로 작성할 수 있습니다.

오브젝트 스코어. 매스, 관성, 기여도를 비롯한 오브젝트 스코어를 표시하며 다음과 같은 옵션이 있습니다.

- 다음 범주 포함. 선택한 분석변수의 범주 표시기를 나타냅니다.
- 오브젝트 스코어 레이블 기준. 레이블변수로 지정된 변수 목록에서 오브젝트를 설명할 변수를 하나 선택할 수 있습니다.

판별 측도. 변수 및 차원별로 판별 측도를 표시합니다.

반복 히스토리. 각 반복에 대하여 설명된 분산, 손실 함수, 설명된 분산의 증가량을 표시합니다.

원 변수간의 상관. 원래 변수의 상관 행렬과 그 행렬의 고유 값을 나타냅니다.

변환된 변수간의 상관. 변환된(최적으로 척도화된) 변수의 상관 행렬과 그 행렬의 고유 값을 나타냅니다.

범주 수량화 및 기여도. 선택된 변수의 각 차원에 대해 매스, 관성, 기여도를 비롯한 범주 수량화(좌표)를 제공합니다.

참고: 매스, 관성을 비롯하여 좌표 및 기여도가 피벗 테이블 결과에서 별도의 레이어에 표시됩니다. 기본적으로 좌표가 표시됩니다. 기여도를 표시하려면 테이블을 활성화(두 번 누르기)하고 레이어 드롭다운 목록에서 기여도를 선택합니다.

기술통계. 선택한 변수의 빈도, 결측값의 수, 최빈값을 표시합니다.

다중 대응일치 분석 저장

저장 대화 상자를 사용하면 이산화 데이터, 오브젝트 스코어 및 변환된 값을 외부 IBM SPSS Statistics 데이터 파일 또는 현재 세션의 데이터 세트에 저장할 수 있습니다. 변환된 값과 오브젝트 스코어를 활성 데이터 세트에 저장할 수도 있습니다.

- 데이터 세트는 현재 세션을 진행하는 동안 사용할 수 있지만 이를 명시적인 데이터 파일로 저장하지 않으면 이후의 세션에서는 사용할 수 없습니다. 데이터 세트 이름은 변수이름 규칙을 준수해야 합니다.
- 파일 이름이나 데이터 세트 이름은 저장된 각 데이터 유형에 따라 서로 달라야 합니다.
- 오브젝트 스코어나 변환된 값을 활성 데이터 세트에 저장하는 경우 다중 명목 차원의 수를 지정할 수 있습니다.

다중 대응일치 분석 대상도표

대상도표 대화 상자를 사용하면 원하는 도표의 유형과 도표화할 변수를 지정할 수 있습니다.

오브젝트 포인트. 오브젝트 포인트 도표가 나타납니다.

대상 및 중심값(Bi-플롯). 변수 중심값에 따라 오브젝트 포인트가 도표화됩니다.

Bi-플롯 변수. Bi-플롯에 대해 모든 변수를 사용하거나 서브세트를 선택할 수 있습니다.

레이블 오브젝트. 오브젝트를 선택한 변수의 범주로 설명하거나(옵션 대화 상자에서 범주 표시기 값 또는 값 레이블을 선택) 케이스의 수로 설명하도록 선택할 수 있습니다. 변수를 선택하면 변수마다 도표 하나가 작성됩니다.

다중 대응일치 분석 변수 도표

변수 도표 대화 상자를 사용하면 원하는 도표 유형과 도표화할 변수를 지정할 수 있습니다.

범주도표. 선택된 각 변수에 대해 중심값 좌표의 도표를 도표화합니다. 범주는 특정 범주의 오브젝트 중심값에 있습니다.

결합범주도표. 선택한 각 변수의 중심값 좌표에 대한 단일 도표입니다.

변환 도표. 최적 범주 수량화와 범주 표시기의 도표를 나타냅니다. 차원 수를 지정할 수 있습니다. 각 차원에 대해 도표 하나가 생성됩니다. 선택된 각 변수에 대해 잔차 도표 표시를 선택할 수도 있습니다.

판별 측도. 선택된 변수에 대해 판별 측도의 단일 도표를 작성합니다.

MULTIPLE CORRESPONDENCE 명령 추가 기능

선택을 구문 창에 붙여넣고 MULTIPLE CORRESPONDENCE 명령 구문을 편집하여 다중 대응일치 분석을 사용자 정의할 수 있습니다. 명령 구문을 사용하여 수행할 수 있는 추가 기능은 다음과 같습니다.

- SAVE 하위 명령을 사용하여 변환된 변수, 오브젝트 스코어 및 근사값의 루트 이름을 지정하고 이를 활성 데이터 세트에 저장할 수 있습니다.
- PLOT 하위 명령을 사용하여 각 도표에 대한 레이블의 최대 길이를 개별적으로 지정할 수 있습니다.
- PLOT 하위 명령을 사용하여 잔차 도표에 대한 개별 변수 목록을 지정할 수 있습니다.

명령 구문에 대한 자세한 내용은 *Command Syntax Reference*를 참조하십시오.

제 7 장 다차원척도법(PROXSCAL)

다차원척도법에서는 오브젝트 간의 근접성 척도 세트 구조를 탐색합니다. 이를 위해 개념적 저차원 공간에 있는 점 사이의 거리가 주어진 유사성(상이성) 척도와 가능한 한 가깝게 일치되도록 이 공간의 특정 위치에 관측 값이 할당됩니다. 그 결과 해당 저차원 공간에 오브젝트의 최소제곱 표현이 얻어지며, 이 표현을 통해 데이터를 더 심도 있게 이해할 수 있습니다.

예제. 다차원척도법은 인지 관계를 결정하는 데 매우 유용합니다. 예를 들어, 제품 이미지를 고려하는 경우에 경쟁사의 제품과 자사 제품 간에 인지되는 유사성(또는 근접성)을 설명하는 데이터 세트를 얻을 수 있도록 시장 설문조사를 할 수 있습니다. 이러한 근접성과 독립변수(예: 가격)를 사용하여 사람들이 이 제품을 어떻게 생각하는지에 대한 중요한 변수가 무엇인지를 결정할 수 있고 그에 따라 이미지를 조정할 수 있습니다.

통계 및 도표. 반복 히스토리, 스트레스 척도, 스트레스 분해, 공통 공간좌표, 최종 설정내의 오브젝트 거리, 개별공간 가중치, 개별공간, 변환된 근접성, 변환 독립변수, 스트레스 도표, 공통 공간 산점도, 개별공간 가중치 산점도, 개별공간 산점도, 변환 도표, Shepard 잔차 도표, 독립변수 변환 도표를 구합니다.

다차원척도법 데이터 고려 사항

데이터. 데이터는 근접 행렬 또는 근접 행렬로 변환되는 변수 형태로 제공될 수 있습니다. 행렬의 형식은 열 또는 한조의 열이 될 수 있습니다. 근접성을 비율, 구간, 순서 또는 스플라인 척도 수준으로 취급할 수 있습니다.

가정. 셋 이상의 변수를 지정해야 합니다. 차원 수는 오브젝트 수에서 1을 뺀 값을 초과할 수 없습니다. 다중 임의 시작과 결합되는 경우에는 차원 수 축소를 생략합니다. 하나의 소스만 지정하는 경우에 모든 모형은 회귀 모형으로 간주됩니다. 따라서 기본값으로 회귀 모형에 대한 분석을 수행합니다.

관련 프로시저. 모든 변수를 연속 수준으로 척도화하는 것은 표준 다차원척도법 분석에 해당합니다.

다차원척도법 수행 방법

1. 메뉴에서 다음을 선택합니다.

분석 > 척도 > 다차원척도법(PROXSCAL)...

이렇게 하면 데이터 형식 대화 상자가 열립니다.

2. 데이터의 형식을 지정합니다.

데이터 형식. 데이터가 근접성 척도로 구성되어 있는지 또는 데이터로부터 근접 행렬을 작성할지를 지정합니다.

소스의 수. 데이터가 근접 행렬인 경우 근접성 척도에 대한 단일 소스 또는 다중 소스를 지정합니다.

단일 소스 근접성에 대한 단일 소스가 있는 경우 데이터 세트를 한 조의 열에서의 근접 행렬로 형식화할지 또는 각 근접성의 행과 열을 구분하는 두 개의 개별변수가 있는 단일 열에서의 근접 행렬로 형식화할지 지정합니다.

- 한조의 열에서의 근접 행렬(*The proximities are in a matrix across columns*). 근접 행렬이 오브젝트 수와 같은 열 수로 산포됩니다. 한조의 열에서의 근접 행렬 대화 상자가 표시됩니다.
- 단일 열에서의 근접 행렬(*The proximities are in a single column*). 근접 행렬은 단일 열 또는 변수로 축소됩니다. 각 셀의 행과 열을 식별하는 2개의 추가 변수가 있어야 합니다. 그러면 단일 열에서의 근접성 대화 상자가 표시됩니다.

다중 소스 근접성에 대한 다중 소스가 있는 경우 데이터 세트를 한 조의 열에서 누적 근접 행렬로 형식화할지, 열마다 소스가 하나인 다중 열에서의 근접 행렬로 형식화할지, 단일 열에서의 근접 행렬로 형식화할지 지정합니다.

- 한조의 열에서의 누적 근접 행렬(*The proximities are in stacked matrices across columns*). 근접 행렬이 오브젝트 수와 동일한 열 수에 걸쳐 산포되고 소스 수만큼의 오브젝트 수와 동일한 행 수에 걸쳐 수직 누적됩니다. 한조의 열에서의 근접 행렬 대화 상자가 표시됩니다.
- 열에서의 근접 행렬, 열마다 1소스(*The proximities are in columns, one source per column*). 근접 행렬이 여러 열 또는 변수로 축소됩니다. 각 셀의 행과 열을 식별하는 2개의 추가 변수가 있어야 합니다. 그러면 열에서의 근접성 대화 상자가 표시됩니다.
- 단일 열에서의 누적 근접 행렬(*The proximities are stacked in a single column*). 근접 행렬은 단일 열 또는 변수로 축소됩니다. 각 셀의 행, 열, 소스를 식별하는 3개의 추가 변수가 있어야 합니다. 그러면 단일 열에서의 근접성 대화 상자가 표시됩니다.

3. 정의를 클릭합니다.

한 조의 열에 있어서의 근접 행렬

데이터 형식 대화 상자에서 단일 소스 또는 다중 소스에 대한 근접 행렬 데이터 모형을 선택한 후 다음을 수행합니다.

1. 셋 이상의 근접변수를 선택합니다. 목록에 있는 변수의 순서와 근접성 열의 순서가 일치하는지 확인합니다.
2. 선택적으로 근접변수의 수와 동일한 가중 변수의 수를 선택합니다. 가중치의 순서와 가중하려는 근접성의 순서가 일치하는지 확인합니다.
3. 다중 소스가 있는 경우 선택적으로 원시변수를 선택합니다. 각 근접변수에 있는 케이스 수는 근접변수의 수와 소스 수를 곱한 값이어야 합니다.

또한 다차원척도법 모형 정의, 공통 공간에 제한 지정, 수렴 기준 설정, 사용할 초기설정 지정, 도표 및 결과 선택 등을 할 수 있습니다.

열에서의 근접 행렬

데이터 형식 대화 상자에서 다중 소스에 대한 다중 열 모형을 선택한 후 다음 중 하나를 수행합니다.

1. 둘 이상의 근접변수를 선택합니다. 각 변수는 개별 소스로부터의 근접 행렬인 것으로 가정합니다.
2. 행 변수를 선택하여 각 근접변수에서 근접성에 대한 행 위치를 정의합니다.
3. 열 변수를 선택하여 각 근접변수에서 근접성에 대한 열 위치를 정의합니다. 행/열을 지정하지 않은 근접 행렬의 셀은 결측으로 취급됩니다.
4. 선택적으로 근접변수의 수와 동일한 가중 변수의 수를 선택합니다.

또한 다차원척도법 모형 정의, 공통 공간에 제한 지정, 수렴 기준 설정, 사용할 초기설정 지정, 도표 및 결과 선택 등을 할 수 있습니다.

단일 열에서의 근접 행렬

데이터 형식 대화 상자에서 단일 소스 또는 다중 소스에 대한 단일 열 모형을 선택한 후 다음을 수행합니다.

1. 근접변수를 선택합니다. 하나 이상의 근접 행렬을 가정합니다.
2. 행 변수를 선택하여 근접변수에서 근접성에 대한 행 위치를 정의합니다.
3. 열 변수를 선택하여 근접변수에서 근접성에 대한 열 위치를 정의합니다.
4. 다중 소스가 있는 경우 원시변수를 선택합니다. 각 소스에 대해 행/열을 지정하지 않은 근접 행렬의 셀은 결측으로 취급됩니다.
5. 선택적으로 가중 변수를 선택합니다.

또한 다차원척도법 모형 정의, 공통 공간에 제한 지정, 수렴 기준 설정, 사용할 초기설정 지정, 도표 및 결과 선택 등을 할 수 있습니다.

데이터로부터 근접 행렬 작성

데이터 형식 대화 상자에서 데이터로부터 근접 행렬을 작성하도록 선택한 후 다음을 수행합니다.

1. 변수 간 거리를 작성하려면(데이터로부터 척도 계산하기 대화 상자 참조) 셋 이상의 변수를 선택합니다. 이들 변수는 근접 행렬(다중 소스인 경우 여러 개의 행렬)을 작성하는 데 사용됩니다. 케이스 간의 거리를 작성하려면 하나의 변수만 필요합니다.
2. 다중 소스가 있는 경우 원시변수를 선택합니다.
3. 선택적으로 근접 행렬 작성에 대한 척도를 선택합니다.

또한 다차원척도법 모형 정의, 공통 공간에 제한 지정, 수렴 기준 설정, 사용할 초기설정 지정, 도표 및 결과 선택 등을 할 수 있습니다.

데이터로부터 척도 계산하기

다차원척도법은 상이성 데이터를 사용하여 척도화 해법을 작성합니다. 데이터가 다변량 데이터(측정된 변수의 여러 값)인 경우 다차원척도화 해법을 계산하려면 상이성 데이터를 작성해야 합니다. 데이터로부터 상이성 척도를 작성하기 위한 세부사항을 지정할 수 있습니다.

척도. 분석에 필요한 상이성 척도를 지정할 수 있습니다. 해당 데이터 유형에 해당하는 척도에서 하나를 선택한 후 척도 유형을 나타내는 드롭다운 목록에서 하나를 선택합니다. 사용 가능한 선택 사항은 다음과 같습니다.

- **구간.** 유클리드 거리, 유클리드 거리 제곱값, Chebychev, 블록, Minkowski 또는 사용자 정의 간격을 선택할 수 있습니다.
- **개수.** 카이제곱 척도 또는 파이제곱 척도를 선택할 수 있습니다.
- **이분형.** 유클리드 거리, 유클리드 거리 제곱값, 크기 차이, 패턴 차이, 분산, Lance와 Williams를 선택할 수 있습니다.

거리 행렬 만들기. 분석 단위를 선택할 수 있습니다. 변수 사이의 거리 또는 케이스들 사이의 거리를 선택합니다.

값 변환. 크게 다른 척도로 변수를 측정하는 경우 등과 같은 특정 상황에서 이분형 데이터에 적용할 수 없는 근접성을 계산하기 전에 값을 표준화해야 할 수도 있습니다. 표준화 드롭다운 목록에서 표준화 방법을 선택합니다. 표준화하지 않아도 되면 지정않음을 선택합니다.

다차원척도법 모형 정의

모형 대화 상자에서 척도화 모형, 차원의 최소값 및 최대값, 근접 행렬의 구조, 근접성에 사용할 변환을 지정할 수 있고 근접성이 각 소스 내에 개별적으로 변환되는지 또는 소스와 무관하게 변환되는지를 지정할 수 있습니다.

척도화 모형. 다음 중 하나를 선택합니다.

- **항등.** 모든 소스에 동일한 설정을 합니다.
- **가중치를 부여한 유클리디안.** 이 모형은 개인차 모형입니다. 각 소스에는 개별공간이 있고 공통 공간의 각 차원은 서로 다르게 가중됩니다.
- **일반화 유클리디안.** 이 모형은 개인차 모형입니다. 각 소스에는 공통 공간의 회전에 해당하는 개별공간이 있고 차원의 서로 다른 가중치가 있습니다.
- **축소계수.** 개별공간의 계수를 지정할 수 있는 일반화 유클리디안 모형입니다. 1 이상이고 차원의 최대 수보다 작은 계수를 지정해야 합니다.

형태. 근접성을 근접 행렬의 하부-삼각 부분에서 또는 상부-삼각 부분에서 취해야 할지를 지정합니다. 완전행렬 사용 여부를 지정할 수 있으며 이 경우 상부-삼각 및 하부-삼각의 부분의 가중 합계를 분석합니다. 모든 경우에 지정한 부분을 사용하더라도 대각 원소를 포함하는 완전행렬을 지정해야 합니다.

근접성. 사용하는 근접 행렬의 유사성 또는 상이성 측도 포함 여부를 지정합니다.

근접성 변환. 다음 중 하나를 선택합니다.

- **비율.** 변환된 근접성은 변환전 근접성에 비례합니다. 양의 값을 갖는 근접성에서만 사용합니다.
- **구간척도.** 변환된 근접성은 변환전 근접성에 절편 항을 더한 값과 비례합니다. 절편의 사용은 모든 변환된 근접성이 양수라는 것을 의미합니다.
- **순서.** 변환된 근접성은 변환전 근접성과 순서가 동일합니다. 동률 근접성을 그 상태로 유지할지 비동률화할지 지정합니다.
- **스플라인.** 변환된 근접성은 변환전 근접성에 대하여 감소하지 않는 평활 조각별 다항 변환입니다. 다항의 차수와 내부 마디 수를 지정합니다.

적용할 변환. 각 소스 내의 근접성만 서로 비교할지 또는 소스와 무관하게 비교할지 지정합니다.

차원. 기본적으로 두 개의 차원(최소 = 2, 최대 = 2)에서 해를 계산합니다. 1부터 시작하여 오브젝트 수에서 1을 뺀 값 사이에 있는 정수 중 최소 및 최대값을 선택합니다. 단, 최소값은 최대값보다 작거나 같아야 합니다. 프로시저에서는 최대 차원의 해를 계산한 다음 최소 차원에 도달할 때까지 단계별로 차원을 축소합니다.

다차원척도법 제한

제한 대화 상자에서 공통 공간에 대한 제한을 둘 수 있습니다.

공통 공간에 대한 제한. 원하는 유형의 제한을 지정합니다.

- **제한 없음.** 공통 공간에 대한 제한을 두지 않습니다.
- **일부 좌표를 고정.** 선택된 첫 번째 변수에는 첫 번째 차원의 오브젝트에 대한 좌표가 포함되고, 두 번째 변수는 두 번째 차원의 좌표에 상응하는 방식입니다. 결측값은 차원의 좌표가 고정되지 않은 것을 나타냅니다. 선택한 변수의 수는 요청한 차원의 최대 수와 동일해야 합니다.
- **독립변수의 선형결합.** 공통 공간을 선택한 변수의 선형결합으로 제한합니다.

제한 변수. 공통 공간에 대한 제한을 정의하는 변수를 선택합니다. 선형 결함을 지정한 경우에는 제한 변수에 대하여 구간, 명목, 순서 또는 스플라인 변환을 지정합니다. 각 경우에 각 변수의 케이스 수는 오브젝트의 수와 동일해야 합니다.

다차원척도법: 옵션

옵션 대화 상자에서 초기설정 유형 선택, 반복 및 수렴 기준 지정, 표준 또는 빠른 알고리즘 선택 등을 수행할 수 있습니다.

초기설정. 다음 선택 사항 중 하나를 선택합니다.

- **심플렉스.** 최대 차원에서 오브젝트는 서로 같은 거리에 놓입니다. 이러한 고차원의 설정을 개선하기 위해 한 번의 반복을 실행합니다. 그리고 나서 모형 대화 상자에서 지정한 차원의 최대 수를 갖는 초기설정을 구하려면 차원 축소 작업을 수행합니다.

- **Torgerson.** 고전적인 척도화 해를 초기설정으로 사용합니다.
- **단일 임의 시작.** 설정을 임의로 선택합니다.
- **다중 임의 시작.** 여러 개의 설정을 임의로 선택하고 가장 작게 정규화된 원시 스트레스를 초기설정으로 사용합니다.
- **사용자 정의.** 사용자 고유의 초기설정 좌표가 있는 변수를 선택합니다. 선택한 변수의 수는 지정한 차원의 최대 수와 동일해야 하며, 차원 1 좌표에 첫 번째 변수가 대응되고 차원 2 좌표에 두 번째 변수가 대응되며 나머지 역시 마찬가지로 대응되어야 합니다. 각 변수에 있는 케이스의 수는 오브젝트의 수와 동일해야 합니다.

반복수 기준. 반복수 기준 값을 지정합니다.

- **스트레스 수렴.** 연속적인 정규화된 원시 스트레스 값의 차이가 여기에서 지정한 0.0과 1.0 사이의 숫자보다 작으면 알고리즘은 반복을 중단합니다.
- **최소 스트레스.** 정규화된 원시 스트레스가 여기에서 지정한 0.0과 1.0 사이의 숫자보다 작으면 알고리즘은 반복을 중단합니다.
- **최대반복수.** 위의 조건이 하나도 충족되지 않으면 여기에서 지정한 수만큼 알고리즘을 통해 반복이 수행됩니다.
- **빠른 알고리즘 사용.** 빠른 알고리즘을 사용하면 알고리즘 속도가 증가됩니다. 이 알고리즘은 회귀 모형을 제외한 다른 모형에서 사용할 수 없으며 제한이 있는 모형에서도 마찬가지로입니다.

다차원척도법 도표, 버전 1

도표 대화 상자에서 어떠한 도표를 작성할 것인지 지정할 수 있습니다. 이 주제에서는 열에서의 근접 행렬 데이터 형식이 있는 경우의 도표 대화 상자에 대해 설명합니다. 개별공간 가중치, 변환전 및 변환된 근접성, 변환된 근접성과 거리 도표에 대해 도표가 작성되는 소스를 지정합니다. 사용할 수 있는 소스 목록은 기본 대화 상자에 있는 근접변수의 목록입니다.

스트레스 도표. 정규화된 원시 스트레스와 차원 간의 도표를 작성합니다. 이 도표는 차원의 최대 수가 차원의 최소 수보다 큰 경우에만 작성됩니다.

공통 공간 도표. 공통 공간의 좌표에 대한 산점도 행렬을 표시합니다.

개별공간 도표. 각 소스에 대하여 개별공간의 좌표를 산점도 행렬에 나타냅니다. 개인차 모형 중 하나를 모형 대화 상자에서 지정한 경우에만 이 도표를 사용할 수 있습니다.

개별공간 가중치. 개별공간 가중치의 산점도를 작성합니다. 개인차 모형 중 하나를 모형 대화 상자에서 지정한 경우에만 이 도표를 사용할 수 있습니다. 가중치를 부여한 유클리디안 모형의 경우 각 축마다 한 차원씩 가중치가 도표에 표시됩니다. 일반화 유클리디안 모형에서는 차원마다 도표를 작성하고 이 도표는 해당 차원의 회전과 가중치를 나타냅니다. 축소계수 모형은 일반화 유클리디안 모형으로 동일한 도표를 작성하지만 개별공간의 차원 수를 축소시킵니다.

변환전 및 변환된 근접성 도표. 변환전 근접성과 변환된 근접성 도표를 작성합니다.

변환된 근접성과 거리 도표. 변환된 근접성과 거리 도표를 작성합니다.

변환된 독립변수. 독립변수에 대한 변환 도표를 작성합니다.

변수와 차원 간의 상관. 독립변수와 공통 공간의 차원 간에 상관도표를 나타냅니다.

다차원척도법 도표, 버전 2

도표 대화 상자에서 어떠한 도표를 작성할 것인지 지정할 수 있습니다. 이 주제에서는 데이터 형식이 열에서의 근접 행렬을 제외한 나머지 형태인 경우의 도표 대화 상자에 대해 설명합니다. 개별공간 가중치, 변환전 및 변환된 근접성, 변환된 근접성과 거리 도표에 대해 도표가 작성되는 소스를 지정합니다. 입력한 소스 수는 기본 대화 상자에서 지정한 원시변수 값이어야 하며 그 범위는 1부터 소스의 수까지입니다.

다차원척도법 결과

결과 대화 상자를 사용하면 표시되는 결과의 양을 조절하고 일부를 개별 파일에 저장할 수 있습니다.

출력. 다음 중 하나 이상을 출력하도록 선택합니다.

- 공통 공간좌표. 공통 공간의 좌표를 표시합니다.
- 개별공간좌표. 개별공간의 좌표를 표시합니다. 단, 모형이 회귀 모형이 아닌 경우에만 사용할 수 있습니다.
- 개별공간 가중치. 개별공간 가중치를 표시합니다. 단, 개별 차이 모형 중 하나를 지정한 경우에만 사용할 수 있습니다. 모형에 따라서 공간 가중치가 회전 가중치와 차원 가중치로 분해되어 나타나기도 합니다.
- 거리. 설정에 있는 오브젝트 간의 거리를 표시합니다.
- 변환된 근접성. 설정에 있는 오브젝트 간의 변환된 근접성을 나타냅니다.
- 입력 데이터. 변환전 근접성을 포함합니다. 데이터 가중치, 초기설정 및 독립변수의 고정 좌표가 있는 경우 이러한 값도 포함합니다.
- 임의시작에 대한 스트레스. 난수 시드와 각 임의 시작에 대한 정규화된 원시 스트레스 값을 표시합니다.
- 반복 히스토리. 기본 알고리즘의 반복 히스토리를 표시합니다.
- 다중 스트레스 측도. 서로 다른 스트레스 값을 표시합니다. 테이블에는 정규화된 원시 스트레스, 스트레스-I, 스트레스-II, S-스트레스, 설명되는 산포(DAF) 및 Tucker의 적합계수 값이 포함됩니다.
- 스트레스 분해. 최종 정규화된 원시 스트레스의 오브젝트 및 소스 분해를 표시합니다. 오브젝트별 평균과 소스별 평균이 포함됩니다.
- 변환된 독립변수. 선형 결합 제한을 선택한 경우 변환된 독립변수와 대응하는 회귀 가중치를 표시합니다.
- 변수와 차원 간의 상관. 선형 결합 제한을 선택한 경우 독립변수와 공통 공간의 차원 간의 상관을 표시합니다.

새 파일로 저장. 공통 공간좌표, 개별공간 가중치, 거리, 변환된 근접성, 변환된 독립변수를 개별 IBM SPSS Statistics 데이터 파일에 저장할 수 있습니다.

PROXSCAL 명령 추가 기능

선택을 구문 창에 붙여넣고 결과로 표시된 PROXSCAL 명령 구문을 편집하면 근접성 분석의 다차원척도법을 사용자 정의할 수 있습니다. 명령 구문을 사용하여 수행할 수 있는 추가 기능은 다음과 같습니다.

- PLOT 하위 명령을 사용하여 변환 및 잔차 도표에 대한 개별 변수 목록을 지정합니다.
- PLOT 하위 명령을 사용하여 개별공간 가중치, 변환, 잔차 도표에 대한 개별 소스 목록을 지정합니다.
- PLOT 하위 명령을 사용하여 표시할 독립변수 변환 도표의 서브세트를 지정합니다.

명령 구문에 대한 자세한 내용은 *Command Syntax Reference*를 참조하십시오.

제 8 장 다차원 확장(PREFSCAL)

다차원 확장법 프로시저는 두 오브젝트 세트 사이의 관계를 시각적으로 조사할 수 있도록 수량화된 공통 척도를 찾는 데 사용됩니다.

예제. 21명의 사람들에게 15가지 아침 식사 메뉴 중 좋아하는 음식에 순서대로 1부터 15까지 순서를 매기도록 요청한 경우를 생각해 봅시다. 다차원 확장법을 사용하면 사람들의 아침 식사 메뉴 선호도가 크게 두 가지 방식에서, 즉 부드러운 빵과 딱딱한 빵, 기름진 음식과 기름기 없는 음식 사이에서 갈리는 것을 확인할 수 있습니다.

또는 운전자 그룹에게 26개의 자동차 모형을 보여 주고 10개 속성에 대해 "1=전혀 닮지 않음"부터 "6=매우 사실적임"까지 6단계 척도로 순위를 매기도록 요청한 경우를 생각해 볼 수 있습니다. 각 응답자에 대해 평균을 구한 값은 유사도로 간주됩니다. 다차원 확장법을 사용하면 유사 모형의 군집을 찾아내고 이들이 가장 밀접하게 관련되어 있는 속성을 찾을 수 있습니다.

통계 및 도표. 다차원 확장법 프로시저를 사용하면 반복 히스토리, 스트레스 측도, 스트레스 분해, 공통 공간좌표, 최종 설정 내의 오브젝트 거리, 개별공간 가중치, 개별공간, 변환된 근접성, 스트레스 도표, 공통 공간 산점도, 개별공간 가중치 산점도, 개별공간 산점도, 변환 도표, Shepard 잔차 도표를 작성할 수 있습니다.

다차원 확장법 데이터 고려 사항

데이터. 데이터는 사각형 근접 행렬의 형태로 제공됩니다. 각 열은 개별 열 오브젝트로 간주됩니다. 근접 행렬의 각 행은 개별 행 오브젝트로 간주됩니다. 근접성에 대한 다중 소스가 있는 경우 행렬이 누적됩니다.

가정. 적어도 두 개 이상의 변수를 지정해야 합니다. 해법의 차원 수는 오브젝트 수에서 1을 뺀 값을 초과할 수 없습니다. 하나의 소스만 지정하는 경우에 모든 모형은 회귀 모형으로 간주됩니다. 따라서 기본값으로 회귀 모형에 대한 분석을 수행합니다.

다차원 확장법 수행 방법

1. 메뉴에서 다음을 선택합니다.

분석 > 척도 > 다차원 확장(PREFSCAL)...

2. 사각형 근접 행렬의 열을 식별하는 변수를 두 개 이상 선택합니다. 각 변수는 개별 열 오브젝트를 나타냅니다.
3. 또는 열 오브젝트변수 개수에 해당하는 수의 가중 변수를 선택합니다. 가중 변수의 순서는 가중치가 주어지는 열 오브젝트의 순서와 일치해야 합니다.
4. 또는 행 변수를 선택합니다. 이 변수의 값(또는 값 레이블)은 결과에서 행 오브젝트를 설명하는 데 사용됩니다.
5. 다중 소스가 있는 경우 선택에 따라서는 원시변수를 선택할 수도 있습니다. 데이터 파일에 있는 케이스 수는 행 오브젝트의 수와 소스 수를 곱한 값이어야 합니다.

또한, 다차원 확장법을 위한 모형을 정의하고, 공통 공간에 제한을 두고, 수렴 기준을 설정하고, 사용할 초기설정 값을 지정하고, 도표와 결과를 선택할 수 있습니다.

다차원 확장법 모형 정의

모형 대화 상자를 사용하면 척도화 모형, 차원의 최소 및 최대 수, 근접 행렬의 구조, 근접성에 사용할 변환을 지정할 수 있고, 근접성이 행에 대해 조건적으로 변환되는지, 소스에 대해 조건적으로 변환되는지 또는 소스와 무관하게 변환되는지를 지정할 수 있습니다.

척도화 모형. 다음 중 하나를 선택합니다.

- **항등.** 모든 소스에 동일한 설정을 합니다.
- **가중치를 부여한 유클리디안.** 이 모형은 개인차 모형입니다. 각 소스에는 개별공간이 있고 공통 공간의 각 차원은 서로 다르게 가중됩니다.
- **일반화 유클리디안.** 이 모형은 개인차 모형입니다. 각 소스에는 공통 공간의 회전에 해당하는 개별공간이 있고 차원의 서로 다른 가중치가 있습니다.

근접성. 사용하는 근접 행렬의 유사성 또는 상이성 측도 포함 여부를 지정합니다.

차원. 기본적으로 두 개의 차원(최소 = 2, 최대 = 2)에서 해를 계산합니다. 1부터 시작하여 오브젝트 수에서 1을 뺀 값 사이에 있는 정수 중 최소 및 최대값을 선택할 수 있습니다. 단, 최소값은 최대값보다 작거나 같아야 합니다. 프로시저에서 최대 차원의 해를 계산한 다음 최소 차원에 도달할 때까지 단계별로 차원을 축소합니다.

근접성 변환. 다음 중 하나를 선택합니다.

- **지정없음.** 근접성을 변환하지 않습니다. 원하는 경우 절편 포함을 선택할 수 있습니다. 이 경우 상수항을 기준으로 근접성을 이동할 수 있습니다.
- **선형.** 변환된 근접성은 변환전 근접에 비례합니다. 즉, 기울기는 변환 함수를 통해 추정되고 절편은 0에 고정됩니다. 이를 비율 변환이라고도 합니다. 원하는 경우 절편 포함을 선택할 수 있습니다. 이 경우 상수항을 기준으로 근접성을 이동할 수 있습니다. 이를 구간 변환이라고도 합니다.
- **스플라인.** 변환된 근접성은 변환전 근접성에 대하여 감소하지 않는 평활 조각별 다항 변환입니다. 다항의 차수와 내부 마디 수를 지정할 수 있습니다. 원하는 경우 절편 포함을 선택할 수 있습니다. 이 경우 상수항을 기준으로 근접성을 이동할 수 있습니다.
- **평활.** 후속 값 사이의 차이를 고려하는 데 따른 제한을 비롯하여 변환된 근접성 순서는 변환전 근접 순서와 같습니다. 그 결과는 "평활 순서" 변환입니다. 동률 근접성을 그 상태로 유지할지 비동률화할지 지정할 수 있습니다.
- **순서.** 변환된 근접성은 변환전 근접성과 순서가 동일합니다. 동률 근접성을 그 상태로 유지할지 비동률화할지 지정할 수 있습니다.

적용할 변환. 각 행 내의 근접성만 서로 비교할지, 각 소스 내의 근접성만 서로 비교할지, 행 또는 소스와 무관하게 비교할지 지정합니다. 즉, 변환을 행별로 수행할지, 소스별로 수행할지 또는 모든 근접성에 대해 한 번에 수행할지 지정합니다.

다차원 확장법 제한

제한 대화 상자에서 공통 공간에 대한 제한을 둘 수 있습니다.

공통 공간에 대한 제한. 공통 공간에서 행 또는 열 오브젝트의 좌표를 고정하도록 선택할 수 있습니다.

행/열 제한 변수. 제한이 포함된 과일을 선택하고 공통 공간에 대한 제한을 정의하는 변수를 선택합니다. 선택된 첫 번째 변수에는 첫 번째 차원의 오브젝트에 대한 좌표가 포함되고, 두 번째 변수는 두 번째 차원의 좌표에 상응하는 방식입니다. 결측값은 차원의 좌표가 고정되지 않은 것을 나타냅니다. 선택한 변수의 수는 요청한 차원의 최대 수와 동일해야 합니다. 각 변수에 대한 케이스 수는 오브젝트 수와 같아야 합니다.

다차원 확장법 옵션

옵션 대화 상자를 사용하면 초기설정 유형을 선택하고, 반복수 및 수렴 기준을 지정하고, 스트레스에 대한 페널티항을 설정할 수 있습니다.

초기설정. 다음 선택 사항 중 하나를 선택합니다.

- **전형.** 완전 대칭 MDS 행렬의 사이 블록(행 사이 및 열 사이의 값)을 보완하기 위해 사각형 근접 행렬이 사용됩니다. 완전한 행렬이 형성되면 초기설정으로 고전적인 척도화 해법이 사용됩니다. 삼각 부등식이나 Spearman 거리를 사용하여 대치를 통해 사이 블록을 채울 수 있습니다.
- **Ross-Cliff.** 이중 중심 및 제곱 근접 행렬에 대한 비정칙 값 분해 결과를 행 및 열 오브젝트에 대한 초기값으로 사용하여 Ross-Cliff를 시작합니다.
- **대응일치.** 행 및 열 스코어의 대칭 정규화와 함께 순서가 뒤바뀐 데이터(상이성 대신 유사성)에 대한 대응일치 분석의 결과를 사용하여 대응일치를 시작합니다.
- **중심값.** 고유 값 분해를 사용하여 설정에 행 오브젝트 위치를 지정하여 프로시저를 시작합니다. 그런 다음 지정된 선택의 중심값에 열 오브젝트의 위치를 지정합니다. 선택 수로는 1에서 근접변수의 수 사이에 있는 양의 정수를 지정합니다.
- **다중 임의 시작.** 무작위로 선택한 여러 초기설정에 대해 해를 계산하고 페널티 스트레스가 가장 낮은 해를 가장 적절한 해법으로 표시합니다.
- **사용자 정의.** 고유한 초기설정의 좌표가 포함된 변수를 선택할 수 있습니다. 선택한 변수의 수는 지정한 차원의 최대 수와 동일해야 하며, 차원 1 좌표에 첫 번째 변수가 대응되고 차원 2 좌표에 두 번째 변수가 대응되며 나머지 역시 마찬가지로 대응되어야 합니다. 각 변수의 케이스 수는 행 및 열 오브젝트의 수를 결합한 것과 같아야 합니다. 행 및 열 좌표는 행 좌표 다음에 열 좌표가 나오도록 누적되어야 합니다.

반복수 기준. 반복수 기준 값을 지정합니다.

- **스트레스 수렴.** 연속적인 페널티 스트레스 값의 상대적인 차이가 여기에서 지정한 값보다 작으면 알고리즘을 통한 반복이 중단됩니다. 값으로는 음수가 아닌 숫자를 지정해야 합니다.
- **최소 스트레스.** 여기에서 지정한 값보다 페널티 스트레스가 낮으면 알고리즘이 중단됩니다. 값으로는 음수가 아닌 숫자를 지정해야 합니다.

- **최대반복수.** 위의 조건이 하나도 충족되지 않으면 여기에서 지정한 수만큼 알고리즘을 통해 반복이 수행됩니다.

페널티항. 변환된 근접성의 변동 계수를 기반으로 페널티항 및 Kruskal 스트레스-I의 곱과 적합도 측도가 같아지도록 페널티 스트레스에 대한 최소화를 시도합니다. 이 제어를 사용하면 페널티항의 강도와 범위를 설정할 수 있습니다.

- **강도.** 강도 모수의 값이 더 작을수록 페널티는 더 강해집니다. 0.0과 1.0 사이의 값을 지정합니다.
- **범위.** 이 모수는 페널티가 활성화되는 지점을 설정합니다. 0.0으로 설정하면 페널티가 활성화되지 않습니다. 이 값을 늘리면 변환된 근접성 중에서 더 큰 변동을 사용하여 해법을 검색합니다. 음수가 아닌 값을 지정합니다.

다차원 확장법 도표

도표 대화 상자에서 어떠한 도표를 작성할 것인지 지정할 수 있습니다.

도표. 다음 도표를 사용할 수 있습니다.

- **다중 시작.** 스트레스와 페널티를 모두 표시하는 페널티 스트레스의 누적된 히스토그램을 표시합니다.
- **초기 공통 공간.** 초기 공통 공간의 좌표에 대한 산점도 행렬을 표시합니다.
- **차원별 스트레스.** 차원과 비교한 페널티 스트레스의 선형 도표를 작성합니다. 이 도표는 차원의 최대 수가 차원의 최소 수보다 큰 경우에만 작성됩니다.
- **최종 공통 공간 도표.** 공통 공간의 좌표에 대한 산점도 행렬을 표시합니다.
- **공간 가중치.** 개별공간 가중치의 산점도를 작성합니다. 개인차 모형 중 하나를 모형 대화 상자에서 지정한 경우에만 이 도표를 사용할 수 있습니다. 가중치를 부여한 유클리디안 모형의 경우 각 축마다 한 차원씩 모든 소스에 대한 가중치가 도표에 표시됩니다. 일반화 유클리디안 모형의 경우 각 소스에 대한 해당 차원의 회전 및 가중치를 모두 나타내는 도표가 차원별로 하나씩 작성됩니다.
- **개별공간 도표.** 각 소스의 개별공간에 대한 좌표의 산점도 행렬을 표시합니다. 개인차 모형 중 하나를 모형 대화 상자에서 지정한 경우에만 이 도표를 사용할 수 있습니다.
- **변환 도표.** 변환된 근접성에 대한 변환전 근접성의 산점도를 작성합니다. 변환 적용 방식에 따라 각 행 또는 소스에 개별 색상이 지정됩니다. 변환에 특별한 조건이 없는 경우 단색이 할당됩니다.
- **Shepard 도표.** 변환된 근접성 및 거리에 대한 변환전 근접성을 표시합니다. 거리는 점으로 표시되고 변환된 근접성은 선으로 표시됩니다. 변환 적용 방식에 따라 각 행 또는 소스에 대한 개별 선이 생성됩니다. 변환에 특별한 조건이 없는 경우 선이 하나만 생성됩니다.
- **적합도 산점도.** 거리에 대한 변환된 근접성의 산점도를 표시합니다. 다중 소스를 지정한 경우 각 소스에 개별 색상이 할당됩니다.
- **잔차 도표.** 잔차(변환된 근접성에서 거리를 뺀 값)에 대한 변환된 근접성의 산점도를 표시합니다. 다중 소스를 지정한 경우 각 소스에 개별 색상이 할당됩니다.

행 오브젝트 유형. 도표에 행 오브젝트가 표시되는 방식을 더 구체적으로 제어할 수 있습니다. 모든 색상을 순환하는 데는 선택적인 색상 변수의 값이 사용됩니다. 가능한 모든 표시를 순환하는 데는 선택적인 마커 변수의 값이 사용됩니다.

소스 도표, 개별공간 도표, 적합도 산점도 및 잔차 도표(소스별로 변환을 적용한 경우 변환 도표 및 Shepard 도표 포함)에 대해 도표를 작성할 소스를 지정할 수 있습니다. 입력한 소스 수는 기본 대화 상자에서 지정한 원시변수 값이어야 하며 그 범위는 1부터 소스의 수까지입니다.

행 도표, 행별로 변환을 적용하는 경우 변환 도표 및 Shepard 도표에 대해 도표를 작성할 행을 지정할 수 있습니다. 입력한 행 수는 1부터 행 수까지의 범위 내에 포함되어야 합니다.

다차원 확장법 결과

결과 대화 상자를 사용하면 표시되는 결과의 양을 조절하고 일부를 개별 파일에 저장할 수 있습니다.

출력. 다음 중 하나 이상을 출력하도록 선택합니다.

- **입력 데이터.** 변환전 근접성을 포함합니다. 데이터 가중치, 초기설정 및 고정 좌표가 있으면 이러한 값도 포함합니다.
- **다중 시작.** 난수 시드와 각 임의의 시작에 대한 페널티 스트레스 값을 표시합니다.
- **초기 데이터.** 초기 공통 공간의 좌표를 표시합니다.
- **반복 히스토리.** 기본 알고리즘의 반복 히스토리를 표시합니다.
- **적합도.** 각기 다른 측도를 표시합니다. 테이블에는 여러 가지 적합도, 적합성 평가, 상관, 변동 및 비감소 측도가 포함됩니다.
- **스트레스 분해.** 행, 열 및 소스 평균과 표준 편차를 비롯하여 페널티 스트레스의 오브젝트, 행 및 소스 분해를 표시합니다.
- **변환된 근접성.** 변환된 근접성을 표시합니다.
- **최종 공통 공간 도표.** 공통 공간의 좌표를 표시합니다.
- **공간 가중치.** 개별공간 가중치를 표시합니다. 이 옵션은 개인차 모형 중 하나를 지정한 경우에만 사용할 수 있습니다. 모형에 따라서 공간 가중치가 회전 가중치와 차원 가중치로 분해되어 나타나기도 합니다.
- **개별공간 도표.** 개별공간의 좌표를 표시합니다. 이 옵션은 개인차 모형 중 하나를 지정한 경우에만 사용할 수 있습니다.
- **적합된 거리.** 설정에 있는 오브젝트 간의 거리를 표시합니다.

새 파일로 저장. 공통 공간 좌표, 개별공간 가중치, 거리 및 변환된 근접성을 별도의 IBM SPSS Statistics 데이터 파일에 저장할 수 있습니다.

PREFSCAL 명령 추가 기능

선택을 구문 창에 붙여넣고 결과로 표시된 PREFSCAL 명령 구문을 편집하면 근접성 분석의 다차원 확장법을 사용자 정의할 수 있습니다. 명령 구문을 사용하여 수행할 수 있는 추가 기능은 다음과 같습니다.

- 다중 소스를 사용할 수 있으면 PLOT 하위 명령으로 개별공간, 적합도 산점도 및 잔차 도표(행렬 조건 변환의 경우 변환 도표 및 Shepard 도표 포함)에 대한 다중 소스 목록을 지정할 수 있습니다.
- 행 조건 변환의 경우 PLOT 하위 명령을 사용하여 변환 도표 및 Shepard 도표에 대한 다중 행 목록을 지정할 수 있습니다.
- INPUT 하위 명령을 사용하여 행 ID 변수 대신 행 수를 지정할 수 있습니다.
- INPUT 하위 명령을 사용하여 소스 ID 변수 대신 소스 수를 지정할 수 있습니다.

명령 구문에 대한 자세한 내용은 *Command Syntax Reference*를 참조하십시오.

주의사항

이 정보는 미국에서 제공되는 제품 및 서비스용으로 작성된 것입니다.

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. 이 책에서 IBM 제품, 프로그램 또는 서비스를 언급했다고 해서 해당 IBM 제품, 프로그램 또는 서비스만을 사용할 수 있다는 것을 의미하지는 않습니다. IBM의 지적 재산을 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.

135-700

서울특별시 강남구 도곡동 467-12, 군인공제회관빌딩

한국 아이.비.엠 주식회사

고객만족센터

전화번호: 080-023-8080

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

Intellectual Property Licensing

Legal and Intellectual Property Law

IBM Japan Ltd.

1623-14, Shimotsuruma, Yamato-shi

Kanagawa 242-8502 Japan

다음 단락은 현지법과 상충하는 영국이나 기타 국가에서는 적용되지 않습니다. 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 비IBM의 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.

(1) 독립적으로 작성된 프로그램과 기타 프로그램(본 프로그램 포함)간의 정보 교환 및 (2) 교환된 정보의 상호 이용을 목적으로 본 프로그램에 관한 정보를 얻고자 하는 라이선스 사용자는 다음 주소로 문의하십시오.

135-700

서울특별시 강남구 도곡동 467-12, 군인공제회관빌딩

한국 아이.비.엠 주식회사

고객만족센터

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 국제 프로그램 라이선스 계약(IPLA) 또는 이와 동등한 계약에 따라 제공한 것입니다.

본 문서에 포함된 모든 성능 데이터는 제한된 환경에서 산출된 것입니다. 따라서 다른 운영 환경에서 얻어진 결과는 상당히 다를 수 있습니다. 일부 성능은 개발 단계의 시스템에서 측정되었을 수 있으므로 이러한 측정치가 일반적으로 사용되고 있는 시스템에서도 동일하게 나타날 것이라고는 보증할 수 없습니다. 또한 일부 성능은 추정을 통해 추측되었을 수도 있으므로 실제 결과는 다를 수 있습니다. 이 책의 사용자는 해당 데이터를 본인의 특정 환경에서 검증해야 합니다.

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

IBM이 제시하는 방향 또는 의도에 관한 모든 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 기업의 이름 및 주소와 유사하더라도 이는 전적으로 우연입니다.

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

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 기업의 이름 및 주소와 유사하더라도 이는 전적으로 우연입니다.

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

© 귀하의 회사명) (연도). 이 코드의 일부는 IBM Corp.의 샘플 프로그램에서 파생됩니다.

© Copyright IBM Corp. _연도_. 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 웹의 『저작권 및 상표 정보』(www.ibm.com/legal/copytrade.shtml)에 있습니다.

Adobe, Adobe 로고, PostScript 및 PostScript 로고는 미국 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

[가]

가중치
비선형 정준 상관 분석 29
가중치를 부여한 유클리디안 모형
다차원 확장 52
개별공간 가중치
다차원 확장 55
다차원척도법 49
개별공간 가중치 도표
다차원 확장 54
다차원척도법 48
개별공간 도표
다차원 확장 54
다차원척도법 48
개별공간 좌표
다차원 확장 55
거리
다차원 확장 55
다차원척도법 49
거리측도
대응일치 분석 33
결측값
다중 대응일치 분석 38
범주형 주성분 분석 21
범주형 회귀분석 15
결합범주도표
다중 대응일치 분석 41
범주형 주성분 분석 24
공간 가중치 도표
다차원 확장 54
공유 도표
다차원 확장 54
공동 공간 도표
다차원 확장 54
다차원척도법 48
공동 공간에 대한 제한
다차원 확장 53
공동 공간좌표
다차원 확장 55
다차원척도법 49
관성
대응일치 분석 34

근접성 변환
다차원 확장 52
기술통계
범주형 회귀분석 17

[나]

능형 회귀분석
범주형 회귀분석 16

[다]

다중 대응일치 분석 37, 39
명령 추가 기능 41
저장할 변수 40
최적적도 수준 38
다중 시작 도표
다차원 확장 54
다중 R
범주형 회귀분석 17
다차원 확장 51
결과 55
공동 공간에 대한 제한 53
도표 51, 54
명령 추가 기능 55
모형 52
옵션 53
통계 51
다차원척도법 43, 44, 45, 46
결과 49
도표 43, 48, 49
명령 추가 기능 50
모형 46
옵션 47
제한 47
통계 43
대응일치 분석 31, 32, 33, 34
도표 31
명령 추가 기능 35
통계 31
도표
다차원척도법 48, 49
대응일치 분석 34
범주형 회귀분석 18

도표 (계속)
비선형 정준 상관 분석 29

[바]

반복 히스토리
다중 대응일치 분석 40
다차원 확장 55
다차원척도법 49
범주형 주성분 분석 23
반복수 기준
다차원 확장 53
다차원척도법 47
범주 수량화
다중 대응일치 분석 40
범주형 주성분 분석 23
범주형 회귀분석 17
비선형 정준 상관 분석 29
범주도표
다중 대응일치 분석 41
범주형 주성분 분석 24
범주형 주성분 분석 19, 22
명령 추가 기능 25
저장할 변수 23
최적적도 수준 20
범주형 회귀분석 13
도표 13
명령 추가 기능 18
저장 17
정규화 16
최적적도 수준 14
통계 13
변수 가중치
다중 대응일치 분석 38
범주형 주성분 분석 20
변환 도표
다중 대응일치 분석 41
다차원 확장 54
다차원척도법 48
범주형 주성분 분석 24
변환된 근접성
다차원 확장 55
다차원척도법 49

- 변환된 독립변수
 - 다차원척도법 49
- 보조 오브젝트
 - 범주형 회귀분석 15
- 분산 분석
 - 범주형 회귀분석 17
- 비선형 정준 상관 분석 27, 28, 29
 - 도표 27
 - 명령 추가 기능 30
 - 통계 27
- 빠른 알고리즘
 - 다차원척도법 47

[사]

- 상관
 - 다차원척도법 49
- 상관 도표
 - 다차원척도법 48
- 상관 행렬
 - 다중 대응일치 분석 40
 - 범주형 주성분 분석 23
- 설명된 분산
 - 범주형 주성분 분석 23
- 성분 로딩
 - 범주형 주성분 분석 23
 - 비선형 정준 상관 분석 29
- 성분 로딩 도표
 - 범주형 주성분 분석 24
- 스트레스 도표
 - 다차원 확장 54
 - 다차원척도법 48
- 스트레스 측도
 - 다차원 확장 55
 - 다차원척도법 49
- 신뢰도 통계
 - 대응일치 분석 34

[아]

- 오브젝트 스코어
 - 다중 대응일치 분석 40
 - 범주형 주성분 분석 23
 - 비선형 정준 상관 분석 29
- 오브젝트 포인트 도표
 - 다중 대응일치 분석 41
 - 범주형 주성분 분석 24

- 이산화
 - 다중 대응일치 분석 38
 - 범주형 주성분 분석 21
 - 범주형 회귀분석 14
- 일반화 유클리디안 모형
 - 다차원 확장 52

[자]

- 잔차 도표
 - 다차원 확장 54
- 적합
 - 비선형 정준 상관 분석 29
- 적합도 산점도
 - 다차원 확장 54
- 정규화
 - 대응일치 분석 33
- 제한
 - 다차원척도법 47
- 중심값
 - 비선형 정준 상관 분석 29

[차]

- 차원
 - 대응일치 분석 33
- 척도화 모형
 - 다차원 확장 52
- 초기 공통 공간 도표
 - 다차원 확장 54
- 초기설정
 - 다차원 확장 53
 - 다차원척도법 47
 - 범주형 회귀분석 15
 - 비선형 정준 상관 분석 29
- 최적척도 수준
 - 다중 대응일치 분석 38
 - 범주형 주성분 분석 20
- 최종 공통 공간 도표
 - 다차원 확장 54

[타]

- 투영된 중심값 도표
 - 범주형 주성분 분석 24

[파]

- 판별 측도
 - 다중 대응일치 분석 40
- 판별 측도 도표
 - 다중 대응일치 분석 41
- 페널티항
 - 다차원 확장 53
- 표준화
 - 대응일치 분석 33

[하]

- 회귀 모형
 - 다차원 확장 52
- 회귀계수
 - 범주형 회귀분석 17

B

- Bi-플롯
 - 다중 대응일치 분석 41
 - 대응일치 분석 34
 - 범주형 주성분 분석 24

E

- elastic net
 - 범주형 회귀분석 16

L

- lasso
 - 범주형 회귀분석 16

P

- PREFSCAL 51

T

- Tri-플롯
 - 범주형 주성분 분석 24

