IBM SPSS Missing Values 23

참고

이 정보와 이 정보가 지원하는 제품을 사용하기 전에, 23 페이지의 『주의사항』의 정보를 읽으십시오.

제품 정보

이 개정판은 새 개정판에서 별도로 명시하지 않는 한, IBM SPSS Statistics의 버전 23, 릴리스 0. 수정사항 0 및 모든 후속 릴리스와 수정에 적용됩니다.

목차

제 1 장 결측값 소개

결측값을 가진 케이스는 중요한 과제에 직면해 있습니다. 일반적인 모형 프로시에서는 분석에서 이러한 케이스를 간단히 무시해 버리기 때문입니다. 결측값이 거의 없고(대략 케이스 총 수의 5% 미만) 이러한 값을 임의로 결측으로 간주할 수 있는 경우, 다시 말해 결측된 값이 다른 값의 영향을 받지 않는 경우에는 일반적으로 목록별 삭제가 상대적으로 "안전"합니다. 결측값 옵션은 목록별 삭제가 적절한 방법한지 여부를 판단하는 데 도움을 주며 목록별 삭제가 없을 경우 결측값을 처리하는 방법을 제공합니다.

결측값 분석 다중 대치 프로시저

결측값 옵션은 결측값 처리에 필요한 두 개의 프로시저 세트를 제공합니다.

- 다중 대치 프로시저는 결측값의 다중 대치를 위해 특별히 고안된 솔루션으로, 결측 데이터의 패턴을 분석합니다. 즉, 여러 버전의 데이터 세트를 생성하며 각 데이터 세트에는 고유한 대치된 값 세트가 있습니다. 통계 분석이 수행될 경우 대치된 모든 데이터 세트의 모수 추정값이 풀링되며 일반적으로 하나의 대치만으로 생성되는 값보다 더 정확합니다.
- 결측값 분석은 결측 데이터 분석을 위한 약간씩 다른 기술 도구 세트(특히, Little의 MCAR 검정)와 다양한 단일 대치법을 제공합니다. 다중 대치는 일반적으로 단일 대치보다 우수한 방법으로 간주되고 있습니다.

결측값 작업

다음과 같은 기본 단계를 통해 결측값 분석을 시작할 수 있습니다.

1. **결측값 검사** 결측값 분석 및 패턴 분석을 사용하여 데이터의 결측값 패턴을 탐색하고 다중 대치가 필요한지 여부를 결정합니다.
2. **결측값 대치** 결측 데이터 값 대치를 사용하여 결측값 대치를 곱합니다.
3. **"완료" 데이터를 분석합니다.** 다중 대치된 데이터를 지원하는 프로시저를 사용합니다. 다중 대치된 데이터 세트 분석 및 이러한 데이터를 지원하는 프로시저 목록에 대한 자세한 정보는 18 페이지의 『다중 대치된 데이터 분석』을 참조하십시오.

제 2 장 결측값 분석

결측값 분석 프로시저로 다음 세 가지 기본 기능을 수행할 수 있습니다.

- 결측값이 있는 위치와 크기를 설명합니다. 결측값의 위치. 결측값의 확장 범위. 여러 케이스에서 대응변수가 결측값을 가지고 있는지 여부. 데이터 값이 극단값인지 여부. 임의로 결측된 값인지 여부.
- 목록별, 대응별, 회귀분석별 또는 EM(기대-최대화) 등 여러 결측값 방법에 대해 평균, 표준 편차, 공분산, 상관 등을 추정합니다. 대응별 방법으로 대응별 완전 결측값 개수도 표시합니다.
- 회귀 방법이나 EM 방법을 사용하여 결측값을 추정값으로 채웁니다. 그러나 보다 정확한 결과를 얻기 위해 일반적으로 다중 대치법을 사용합니다.

결측값 분석을 통해 불완전한 데이터로 인해 발생하는 몇 가지 문제를 파악할 수 있습니다. 결측값이 있는 케이스가 결측값이 없는 케이스와 분류상 다른 경우 결과가 잘못될 수 있습니다. 또한 결측 데이터는 원래 의도한 것보다 정보가 부족하므로 계산된 통계가 정확하지 않을 수도 있습니다. 여러 통계 프로시저의 가정은 케이스가 완전하다는 전제 하에 있으므로 결측값으로 인해 필요한 이론이 복잡해질 수 있습니다.

예제. 백혈병 치료 평가에서 여러 가지 변수가 측정되었습니다. 그러나 모든 치료법을 모든 환자에 대해 적용할 수는 없습니다. 결측 데이터 패턴은 출력되거나 표로 만들어지고 무작위로 사용됩니다. EM 분석은 평균, 공분산, 상관 등을 추정하는 데 사용됩니다. 또한 데이터가 임의로 완전히 결측되었는지 파악하는 데 사용됩니다. 결측값은 대치된 값으로 바뀌고 새 데이터 파일에 저장되어 다음 분석에서 사용됩니다.

통계. 비결측값의 수, 평균, 표준 편차, 결측값 수, 극단값 수 등이 포함된 일변량 통계량을 구할 수 있습니다. 목록별, 대응별, EM, 회귀분석 등의 방법을 사용한 평균, 공분산 행렬, 상관 행렬 추정할 수 있습니다. EM 결과를 사용한 Little의 MCAR 검정을 구할 수 있습니다. 다양한 방법을 사용한 평균 요약을 구할 수 있습니다. 결측값과 비결측값을 비교하여 정의한 그룹에 대해: *t* 검정. 변수별 케이스 표시 결측값 패턴 통계를 구합니다.

데이터 고려 사항

데이터. 데이터는 범주형이거나 양적변수(척도 또는 연속형)여야 합니다. 하지만 양적변수에 대해서만 통계를 추정하고 결측 데이터를 대치할 수 있습니다. 각 변수에서 시스템-결측값으로 코딩되지 않은 결측값은 사용자 결측값으로 정의해야 합니다. 예를 들어, 어떤 질문지 항목에 대한 반응이 모름일 때 5로 코딩하여 이를 결측값으로 처리하려면 해당 항목을 사용자 결측값인 5로 코딩해야 합니다.

빈도 가중치. 빈도 (복제) 가중치는 이 프로시저에서 사용됩니다. 복제 가중치가 음수 또는 값이 0인 케이스는 무시됩니다. 정수가 아닌 가중치는 잘립니다.

가정. 목록별, 대응별, 회귀분석별 추정은 결측값 패턴이 데이터 값에 따라 달라지지 않는다는 가정을 따릅니다. 이러한 조건은 **임의로 완전 결측** 또는 MCAR이라고 합니다. 따라서 데이터가 MCAR일 때 추정에 사용되는 모든 방법(EM 방법 포함)은 일관적이며 비편향적인 상관 및 공분산 추정값을 제공합니다. MCAR 가정을 위반하면 목록별, 대응별, 회귀분석별 방법에서 편향적인 추정값이 생성될 수 있습니다. 데이터가 MCAR이 아닐 경우 EM 추정을 사용해야 합니다.

EM 추정은 결측 데이터 패턴이 관측 데이터에만 관련되어 있다는 가정을 따릅니다. 이러한 조건을 **임의로 결측** 또는 MAR이라고 합니다. 이러한 가정을 통해 추정값이 가능한 정보를 사용하여 조정되도록 할 수 있습니다. 예를 들어, 교육과 수입에 대한 연구에서 교육이 낮은 개체가 수입 결측값이 더 많을 수 있습니다. 이 경우 데이터는 MCAR이 아닌 MAR입니다. 즉, MAR의 경우 수입이 기록되는 확률은 개체의 교육 수준에 따라 다릅니다. 이 확률은 교육에 따라 다를 수 있지만 *교육 수준 내의* 수입에 따라 변하지 않습니다. 각 교육 수준 내에서 수입값에 따라 수입이 기록되는 확률도 달라지는 경우(예: 고수입자가 이를 보고하지 않는 경우) 데이터는 MCAR 또는 MAR이 아닙니다. 이는 일반적인 상황이 아니지만 이 경우 두 방법 모두 적합하지 않습니다.

관련 프로시저. 여러 프로시저로 목록별 추정이나 대응별 추정을 사용할 수 있습니다. 선형회귀 모형과 요인분석에서는 결측값이 평균값으로 바뀝니다. 예측 추가 기능 모듈에서는 여러 방법을 사용하여 시계열 분석에서 결측값을 바꿀 수 있습니다.

결측값 분석 구하는 방법

1. 메뉴에서 다음을 선택합니다.

   분석 > 결측값 분석...

2. 통계를 추정하고 선택적으로 결측값을 대치할 하나 이상의 양적(척도)변수를 선택합니다.

선택적으로 다음을 수행할 수 있습니다.

- 범주형 변수(숫자 또는 문자)를 선택하고 범주 수 한계(**최대 범주 수**)를 입력합니다.
- **패턴**을 클릭하여 결측값 패턴을 표로 만듭니다. 자세한 정보는 『결측값 패턴 표시』 주제를 참조하십시오.
- **기술통계**를 클릭하여 결측값의 기술통계를 표시합니다. 자세한 정보는 6 페이지의 『결측값의 기술통계 표시』 주제를 참조하십시오.
- 통계(평균, 공분산 및 상관)를 추정하고 결측값을 대치할 수 있는 방법을 선택합니다. 자세한 정보는 7 페이지의 『통계 추정 및 결측값 대치』 주제를 참조하십시오.
- EM이나 회귀분석을 선택한 경우에는 **변수**를 클릭하여 추정에 사용할 변수 세트를 지정합니다. 자세한 정보는 9 페이지의 『예측 및 예측변수』 주제를 참조하십시오.
- 케이스 레이블 변수를 선택합니다. 이 변수는 개별 케이스를 표시하는 패턴표의 케이스를 설명하는 데 사용됩니다.

결측값 패턴 표시

결측 데이터의 패턴 및 범위를 보여주는 다양한 테이블을 표시할 수 있습니다. 이러한 테이블은 다음을 식별하는 데 도움이 됩니다.

- 결측값의 위치
- 대응변수가 개별 케이스에서 결측값을 갖는지 여부
- 데이터 값의 극단성

표시

세 가지 유형의 테이블을 사용하여 결측값 패턴을 표시할 수 있습니다.

케이스 표 작성. 분석변수의 결측값 패턴은 각 패턴에 대한 빈도 분석과 함께 표 작성 형식으로 표시됩니다. **결측값 패턴에 따라 변수 정렬**을 사용하여 개수 및 변수를 패턴의 유사성에 따라 정렬할지 여부를 지정합니다. **전체 케이스 중 n개 미만인 패턴들 생략**을 사용하여 자주 나타나지 않는 패턴은 삭제합니다.

결측값을 가지는 케이스. 결측값 또는 극단값이 있는 각 케이스는 각 분석변수에 대해 표로 만들었습니다. **결측값 패턴에 따라 변수 정렬**을 사용하여 개수 및 변수를 패턴의 유사성에 따라 정렬할지 여부를 지정합니다.

모든 케이스. 각 케이스는 표로 만들었으며 각 변수에 대한 결측값 및 극단값이 표시됩니다. 케이스는 변수를 **정렬 기준**으로 따로 지정하지 않는 한 데이터 파일에 표시되는 순서대로 나열됩니다.

개별 케이스가 표시되는 테이블에는 다음 기호가 사용됩니다.

+. 상한 극단값

-. 하한 극단값

S. 시스템 결측값

A. 첫번째 유형의 사용자 결측값

B. 두 번째 유형의 사용자 결측값

C. 세 번째 유형의 사용자 결측값

변수

분석에 포함되는 변수의 추가 정보를 표시할 수 있습니다. **추가 정보**를 추가할 변수는 결측 패턴 테이블에 개별적으로 표시됩니다. 양적(척도)변수의 경우 평균을 볼 수 있고 범주형 변수의 경우 각 범주마다 패턴이 있는 케이스 수를 볼 수 있습니다.

- **정렬기준.** 지정한 변수 값의 오름차순이나 내림차순에 따라 케이스가 나열됩니다. **모든 케이스**인 경우에만 사용할 수 있습니다.

결측값 패턴 표시 방법

1. 기본 결측값 분석 대화 상자에서 결측값 패턴을 표시할 변수를 선택합니다.
2. **패턴**을 클릭합니다.
3. 표시할 패턴 테이블을 선택합니다.

결측값의 기술통계 표시

일변량 통계량

일변량 통계량을 사용하여 결측값의 일반적인 범위를 식별할 수 있습니다. 각 변수에 대해 다음이 표시됩니다.

- 비결측값 수
- 결측값의 수 및 백분율

양적(척도)변수에 대해 다음이 표시됩니다.

- 평균
- 표준 편차
- 상한 극단값 및 하한 극단값 수

지시변수 통계

각 변수에 대해 지시변수가 생성됩니다. 이 범주형 변수는 개별 케이스에 대해 변수가 존재하는지 결측되어 있는지를 나타냅니다. 지시변수는 불일치, *t* 검정, 빈도표 작성에 사용됩니다.

퍼센트 불일치. 각 대응변수의 경우, 한 변수에 결측값이 있고 다른 변수에는 비결측값이 있는 케이스의 백분율을 표시합니다. 해당 테이블의 각 대각 원소에는 단일 변수에 대한 결측값 백분율이 포함되어 있습니다.

지시변수에 의해 형성된 그룹의 T 검정. 스튜던트 *t* 통계를 사용하여 두 그룹의 평균을 각 양적변수에 대해 비교합니다. 이 그룹은 변수가 존재하는지 결측되어 있는지를 지정합니다. *t* 통계, 자유도, 결측값 및 비결측값의 개수, 두 그룹의 평균을 표시합니다. *t* 통계와 연관된 양쪽 확률을 표시할 수도 있습니다. 분석을 통해 둘 이상의 검정이 생성되는 경우 이러한 확률을 유의수준 검정에 사용하지 마십시오. 확률은 단일 검정을 계산할 때만 적용됩니다.

범주형 및 지시변수의 교차 분석표. 각 범주형 변수마다 테이블이 표시됩니다. 각 범주에 대해 테이블에는 다른 변수에 대한 비결측값의 빈도와 백분율이 나타납니다. 각 결측값 유형의 백분율도 표시됩니다.

전체 케이스 중 n개 미만으로 결측된 변수 생략. 테이블 크기를 줄이려면 작은 케이스 수만을 계산한 통계를 생략합니다.

기술통계 표시 방법

1. 기본 결측값 분석 대화 상자에서 결측값 기술통계를 표시할 변수를 선택합니다.
2. **기술통계**를 클릭합니다.
3. 표시할 기술통계를 선택합니다.

통계 추정 및 결측값 대치

목록별(완전 케이스만), 대응별, EM(기대-최대화) 및/또는 회귀분석 방법을 사용하여 평균, 표준 편차, 공분산, 상관 등을 추정합니다. 결측값을 대치하도록 선택할 수도 있습니다(복원값 추정). 다중 대치는 일반적으로 결측값 문제 해결에 단일 대치보다 뛰어난 방법으로 간주됩니다. Little의 MCAR 검정은 대치의 필요성을 결정하는 데 여전히 유용한 방법입니다.

목록별 방법

이 방법에서는 완전 케이스만 사용합니다. 분석변수에 결측값이 있는 경우 해당 케이스가 계산에서 생략됩니다.

대응별 방법

이 방법은 대응 분석변수를 확인하며 두 변수에 대해 비결측값이 있는 경우에만 케이스를 사용합니다. 각 대응 변수에 대해 빈도, 평균 및 표준 편차가 별도로 계산됩니다. 케이스의 기타 결측값은 무시되므로 두 변수의 상관 및 공분산은 다른 변수의 결측값에 따라 달라지지 않습니다.

EM 방법

이 방법에서는 부분적인 결측 데이터에 대한 분포를 가정하고 해당 분포에 있는 우도의 추론값을 기준으로 합니다. 각 반복은 E 단계와 M 단계로 구성됩니다. E 단계는 모수의 현재 추정값 및 관측값이 지정되면 "결측" 데이터의 조건 기대를 찾아냅니다. 그런 다음 이러한 기대는 "결측" 데이터를 대체합니다. M 단계에서는 결측 데이터가 채워진 것처럼 모수의 최대 우도 추정값이 계산됩니다. 결측 데이터는 직접 채워지지 않으므로 "결측"이 따옴표로 묶입니다. 대신, 결측 함수가 로그-우도에 사용됩니다.

해당 값이 임의로 완전히 결측되었는지 검정하기 위한 Roderick J. A. Little의 카이제곱 통계가 EM 행렬에 꼬리말로 표시됩니다. 이 검정의 경우, 귀무가설은 데이터가 임의로 완전히 결측되었고 p 값이 0.05 수준에서 유의하다는 것입니다. 값이 0.05보다 작으면 데이터가 임의로 완전히 결측되지 않은 것입니다. 데이터가 임의로 결측(MAR) 또는 임의로 결측되지 않은(NMAR) 상태일 수 있습니다. 둘 중 하나를 가정할 수 없으며 데이터를 분석하여 데이터가 결측되는 방법을 파악해야 합니다.

회귀분석방법

이 방법은 다중 선형회귀 추정값을 계산하며 무작위 성분으로 추정값을 증가시킬 수 있는 옵션을 사용할 수 있습니다. 프로시저에서 각 예측값에 임의로 선택한 완전 케이스의 잔차, 임의 정규편차 또는 t 분포의 임의 편차(잔차 평균 제곱의 제곱근으로 척도화됨)를 추가할 수 있습니다.

EM 추정 옵션

EM 방법에서는 반복적인 프로세스를 사용하여 결측값이 있는 양적(척도)변수의 평균, 공분산 행렬, 상관 등을 추정할 수 있습니다.

분포. EM에서는 지정된 분포에 있는 우도를 기준으로 추론합니다. 기본적으로 정규 분포를 따릅니다. 분포의 꼬리가 정규 분포 꼬리보다 길 경우, 프로시저에서 n 자유도를 사용하여 스튜던트 t 분포의 우도 함수를 구성

하도록 요청할 수 있습니다. 혼합된 정규 분포도 꼬리가 긴 분포를 제공합니다. 혼합된 정규 분포의 표준 편차 비율과 두 분포의 혼합 비율을 지정합니다. 혼합된 정규 분포에서는 분포의 표준 편차만 다르다고 가정합니다. 평균은 같아야 합니다.

최대반복수. 최대 반복 수를 설정하여 정확한 공분산을 추정합니다. 추정값이 수렴되지 않는 경우에도 이 최대 반복수에 도달하면 프로시저가 중단됩니다.

완전한 데이터 저장. 결측값 대신 대치된 값이 있는 데이터 세트를 저장할 수 있습니다. 하지만 공분산을 기반으로 하는 통계에서 대치된 값을 사용하면 해당 모수값을 과소추정하게 됩니다. 과소추정도는 공동으로 관측되지 않는 케이스 수에 비례합니다.

EM 옵션 지정 방법

1. 기본 결측값 분석 대화 상자에서 EM 방법을 사용하여 결측값을 추정할 변수를 선택합니다.
2. 추정에서 **EM**을 선택합니다.
3. 예측 또는 예측변수를 지정하려면 **변수**를 클릭합니다. 자세한 정보는 9 페이지의 『예측 및 예측변수』 주제를 참조하십시오.
4. **EM**을 클릭합니다.
5. 원하는 EM 옵션을 선택합니다.

회귀분석 추정 옵션

회귀분석 방법은 다중 선형회귀 모형을 사용하여 결측값을 추정합니다. 예측된 변수의 평균, 공분산 행렬, 상관 행렬 등이 표시됩니다.

추정 조정. 회귀분석 방법에서는 무작위 성분을 회귀 추정에 추가할 수 있습니다. 잔차, 정규변량, 스튜던트 *t* 변량, 지정않음 등을 선택할 수 있습니다.

- *잔차(Residuals)*. 오차항은 회귀 추정으로 추가될 완전한 케이스의 관측 잔차로부터 무작위로 선택됩니다.
- *정규 변량(Normal Variates)*. 오차항은 기대값 0을 가지며 표준 편차가 회귀분석의 평균 제곱 오차항의 제곱근과 동일한 분포로부터 임의로 추출됩니다.
- *스튜던트 T 변량(Student's t Variates)*. 오차항은 t 분포에서 지정된 자유도로 무작위로 추출되었으며 평균 제곱 오차(RMSE)에 의해 측정됩니다.

예측변수의 최대수. 추정 과정에 사용된 예측(독립) 변수의 최대 제한 수를 설정합니다.

완전한 데이터 저장. 현재 세션의 데이터 세트나 외부 형식 IBM® SPSS® Statistics 데이터 파일을 작성하며 이때 결측값은 회귀분석 방법을 사용하여 추정된 값으로 바뀝니다.

회귀분석 옵션 지정 방법

1. 기본 결측값 분석 대화 상자에서 회귀분석 방법을 사용하여 결측값을 추정할 변수를 선택합니다.
2. 추정에서 **회귀분석**을 선택합니다.
3. 예측 또는 예측변수를 지정하려면 **변수**를 클릭합니다. 자세한 정보는 9 페이지의 『예측 및 예측변수』 주제를 참조하십시오.

4. **회귀분석**을 클릭합니다.
5. 원하는 회귀분석 옵션을 선택합니다.

예측 및 예측변수

기본적으로 모든 양적변수는 EM 및 회귀분석 추정에 사용됩니다. 필요한 경우 추정에서 특정 변수를 예측 및 예측변수로 지정할 수 있습니다. 지정된 변수는 두 목록에 사용할 수 있지만 변수 사용을 제한해야 하는 경우가 있습니다. 예를 들어, 일부 분석가는 결과변수의 값을 추정하는 것을 싫어할 수 있습니다. 여러 추정에 다양한 변수를 사용하여 프로시저를 여러 번 실행할 수 있습니다. 예를 들어, 간호사 등급 항목 세트와 의사 등급 항목 세트가 있는 경우 간호사 항목을 사용하는 프로시저를 실행하여 결측 간호사 항목을 추정하고 다른 프로시저를 실행하여 의사 항목을 추정할 수 있습니다.

회귀분석 방법을 사용할 때는 다른 사항을 고려해야 합니다. 다중 회귀분석에서 대량의 독립변수 서브세트를 사용하면 소량의 서브세트를 사용할 때보다 예측값이 부적합할 수 있습니다. 따라서 변수는 진입 *F*한도를 4.0으로 달성하여 사용되도록 해야 합니다. 이 한도는 구문을 사용하여 변경할 수 있습니다.

예측 및 예측변수 지정 방법

1. 기본 결측값 분석 대화 상자에서 회귀분석 방법을 사용하여 결측값을 추정할 변수를 선택합니다.
2. 추정에서 **EM**이나 **회귀분석**을 선택합니다.
3. **변수**를 클릭합니다.
4. 모든 변수가 아닌 특정 변수를 예측 및 예측변수로 사용하려는 경우 **변수 선택**을 선택하고 변수를 적절한 목록으로 이동합니다.

MVA 명령 추가 기능

명령 구문을 사용하여 수행할 수 있는 추가 기능은 다음과 같습니다.

- MPATTERN, DPATTERN, TPATTERN 등의 하위 명령에서 DESCRIBE 키워드를 사용하여 결측값 패턴, 데이터 패턴, 표로 작성한 패턴 등에 대해 구체적인 변수를 각각 지정합니다.
- DPATTERN 하위 명령으로 데이터 패턴 테이블에 대해 여러 정렬변수를 지정합니다.
- DPATTERN 하위 명령으로 데이터 패턴에 대해 여러 정렬변수를 지정합니다.
- EM 하위 명령으로 허용 오차와 수렴을 지정합니다.
- REGRESSION 하위 명령을 사용하여 허용 오차 및 진입 *F*를 지정합니다.
- EM 하위 명령과 REGRESSION 하위 명령으로 EM 및 회귀에 대해 서로 다른 변수 목록을 지정합니다.
- TTESTS, TABULATE, MISMATCH마다 표시된 케이스를 출력하지 않기 위해 서로 다른 백분율을 지정합니다.

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

제 3 장 다중 대치

다중 대치는 결측값에 대해 가능한 값을 생성하고 이에 따라 일부 "완벽한" 데이터 세트를 만드는 데 사용됩니다. 다중 대치된 데이터 세트를 사용하는 분석 프로시저에서는 각각의 "완벽한" 데이터 세트와 원래 데이터 세트에 결측값이 없는 경우 결과를 추정하는 "완벽한" 풀링 결과를 생성합니다. 일반적으로 이러한 풀링된 결과는 단일 대치법으로 생성되는 결과에 비해 더 정확합니다.

다중 대치된 데이터 고려 사항

분석 변수. 분석 변수는 다음과 같습니다.

- *명목(Nominal).* 변수의 값이 고유한 순위가 없는 범주를 나타내는 경우 해당 변수는 명목으로 취급될 수 있습니다. 예를 들어, 직원이 근무하는 회사의 부서가 있습니다. 종교, 우편번호 또는 종교 단체 등이 명목 변수에 해당합니다.
- *순서(Ordinal).* 변수의 값이 고유한 순위가 있는 범주를 나타내는 경우 해당 변수는 순서로 취급될 수 있습니다. 예를 들어, 매우 불만족에서 매우 만족에 이르는 서비스 만족도 수준이 있습니다. 순서변수의 예로는 만족도나 신뢰도를 나타내는 태도 스코어 및 선호도 등급 스코어가 있습니다.
- *척도(Scale).* 해당 값이 의미 있는 메트릭으로 순서가 지정된 범주를 나타내므로 값 간 거리 비교가 적합한 경우 해당 변수는 척도(연속형)로 처리할 수 있습니다. 척도변수의 예로는 연령과 수입이 있습니다.

  소스 변수 목록에서 변수를 마우스 오른쪽 단추로 클릭하고 팝업 메뉴에서 측정 수준을 선택하여 변수에 대한 측정 수준을 임시로 변경할 수 있지만 프로시저는 적절한 측정 수준이 모든 변수에 지정되었다고 가정합니다. 변수의 측정 수준을 영구적으로 변경하려면

변수 목록의 각 변수 옆에 있는 아이콘은 측정 수준 및 변수 유형을 식별합니다.

표 1. 측정 수준 아이콘

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

빈도 가중치. 빈도 (복제) 가중치는 이 프로시저에서 사용됩니다. 복제 가중치가 음수 또는 값이 0인 케이스는 무시됩니다. 정수가 아닌 가중치는 가장 가까운 정수로 반올림됩니다.

분석 가중치. 분석(회귀분석 또는 표본추출) 가중치는 결측값 요약 및 적합 대치 모형에 통합됩니다. 분석 가중치가 음수이거나 0인 케이스는 제외됩니다.

복합 표본. 다중 대치 프로시저는 분석 가중 변수의 형태로 최종 표본추출 가중치를 받아들일 수 있지만 계층, 군집 또는 기타 복잡한 표본추출 구조를 명시적으로 처리하지 않습니다. 또한 복합 표본추출 프로시저는 현재 다중으로 대치된 데이터 세트를 자동으로 분석하지 않습니다. 풀링을 지원하는 프로시저의 전체 목록은 18 페이지의 『다중 대치된 데이터 분석』을 참조하십시오.

결측값. 사용자 결측값 및 시스템 결측값은 유효한 값으로 처리됩니다. 즉, 값이 대치되고 두 결측값이 대치 모형에서 예측변수로 사용되는 변수의 유효한 값으로 처리될 경우 두 결측값 유형이 대체됩니다. 사용자 결측값 및 시스템 결측값은 또한 결측 분석에서 결측값으로 처리됩니다.

결과 복제(결측 데이터 값 대치). 대치 결과를 정확히 복제하려면 임의 수 생성기와 동일한 초기값, 동일한 데이터 순서, 동일한 변수 순서, 동일한 프로시저 설정을 사용합니다.

- **난수 생성기.** 이 프로시저에서는 대치된 값 계산 과정에서 난수를 생성합니다. 동일한 임의화 결과를 생성하려면 각 대치 결측 데이터 값 프로시저를 실행하기 전에 난수 생성기와 동일한 초기값을 사용합니다.
- **케이스 순서.** 값이 테이스 순서대로 대치됩니다.
- **변수 순서.** 전체 조건 지정 사항(FCS) 대치법은 분석 변수 목록에 지정된 순서에 따라 값을 대치합니다.

다중 대치 전용 대화 상자가 두 개 있습니다.

- 패턴 분석은 데이터의 결측값 패턴에 대한 기술 측도를 제공하며 대치 전 예비 단계로 유용하게 사용할 수 있습니다.
- 결측 데이터 값 대치는 다중 대치 생성에 사용합니다. 완성된 데이터 세트는 다중 대치된 데이터 세트가 지원하는 프로시저로 분석할 수 있습니다. 다중 대치된 데이터 세트 분석 및 이러한 데이터를 지원하는 프로시저 목록에 대한 자세한 정보는 18 페이지의 『다중 대치된 데이터 분석』을 참조하십시오.

패턴(다중 대치)

패턴 분석은 데이터의 결측값 패턴에 대한 기술 측도를 제공하며 대치 전 예비 단계로 유용하게 사용할 수 있습니다.

예제. 한 통신 제공업체가 고객 데이터베이스에서 서비스 사용 패턴을 더 잘 이해하고 싶어합니다. 이 회사는 고객이 사용하는 서비스의 완전한 데이터를 갖고 있지만 회사에서 수집하는 인구 통계학적 정보에 많은 결측값이 있습니다. 결측값의 패턴을 분석하면 다음 단계의 대치를 결정할 수 있습니다. 자세한 정보는 주제를 참조하십시오.

메뉴에서 다음을 선택합니다.

분석 > 다중 대치 > 패턴 분석...

1. 두 개 이상의 분석 변수를 선택합니다. 이 프로시저에서는 이러한 변수에 대한 결측 데이터 패턴을 분석합니다.

선택적 설정

분석 가중치. 이 변수에는 분석(회귀분석 또는 표본추출) 가중치가 포함됩니다. 이 프로시저에서는 결측값 요약의 분석 가중치를 통합합니다. 분석 가중치가 음수이거나 0인 케이스는 제외됩니다.

결과. 다음과 같은 선택적 결과를 사용할 수 있습니다.

- **결측값 요약.** 분석 변수의 수 및 퍼센트, 케이스 또는 하나 이상의 결측값을 가진 개별 데이터 값을 보여주는 패널화된 원도표가 표시됩니다.
- **결측값 패턴.** 표 형식의 결측값 패턴이 표시됩니다. 각 패턴은 분석 변수에 대해 동일한 미완성 데이터 패턴 및 완성 데이터 패턴을 가진 케이스 그룹에 해당합니다. 이 출력결과를 사용하여 데이터에 단조로운 대치법을 사용하거나, 또는 그렇지 않을 경우 데이터에 얼마나 단조로운 패턴을 사용할지를 결정할 수 있습니다. 이 프로시저에서는 분석 변수의 순서를 지정해 단조로운 패턴을 나타내거나 이와 가깝게 접근합니다. 순서를 재지정한 후 단조로운 패턴이 없는 경우에는 분석 변수의 순서가 지정되었을 때 데이터가 단조로운 패턴을 갖는 것으로 결론지을 수 있습니다.
- **가장 높은 빈도의 결측값을 가진 변수.** 내림차순의 결측값(%) 기준으로 정렬된 분석 변수 테이블이 표시됩니다. 테이블에는 척도변수에 대한 기술 통계(평균 및 표준 편차)가 포함됩니다.

  최대 변수 수를 제어하여 화면에 표시되는 변수에 대한 최소 결측값(%)을 표시할 수 있습니다. 두 기준에 부합하는 변수 세트가 표시됩니다. 예를 들어, 최대 변수 수를 50, 최소 결측값(%)을 25로 설정하면 25% 이상의 결측값을 가진 50개의 변수가 테이블에 표시됩니다. 60개의 분석 변수가 있고 그 중 15개만 25% 이상의 결측값을 가진 경우 결과에는 15개의 변수만 포함됩니다.

결측 데이터 값 대치

결측 데이터 값 대치는 다중 대치 생성에 사용합니다. 완성된 데이터 세트는 다중 대치된 데이터 세트가 지원하는 프로시저로 분석할 수 있습니다. 다중 대치된 데이터 세트 분석 및 이러한 데이터를 지원하는 프로시저 목록에 대한 자세한 정보는 18 페이지의 『다중 대치된 데이터 분석』을 참조하십시오.

예제. 한 통신 제공업체가 고객 데이터베이스에서 서비스 사용 패턴을 더 잘 이해하고 싶어합니다. 이 회사는 고객이 사용하는 서비스의 완전한 데이터를 갖고 있지만 회사에서 수집하는 인구 통계학적 정보에 많은 결측값이 있습니다. 게다가 이러한 값은 완전히 무작위로 결측된 것이므로 데이터 세트를 완성하기 위해 여러 대치가 사용됩니다. 자세한 정보는 주제를 참조하십시오.

메뉴에서 다음을 선택합니다.

분석 > 다중 대치 > 결측 데이터 값 대치...

1. 대치 모형에서 두 개 이상의 변수를 선택합니다. 이 프로시저에서는 이러한 변수에 대한 결측 데이터 값의 여러 변수를 대치합니다.
2. 계산할 대치 수를 지정합니다. 기본적으로 이 값은 5입니다.
3. 데이터베이스 또는 대치된 데이터가 기록될 IBM SPSS Statistics 형식의 데이터 세트를 지정합니다.

   결과 데이터 세트는 각 대치를 위해 대치된 값을 갖는 케이스 세트와 더불어 결측 데이터를 가진 원본 케이스 데이터로 구성됩니다. 예를 들어, 원래 데이터 세트에 100개의 케이스가 있고 5개의 대치가 있는 경

우 결과 데이터 세트에는 600개의 케이스가 있게 됩니다. 입력 데이터 세트에 있는 모든 변수는 결과 데이터 세트에 포함됩니다. 기존 변수의 사전 특성(이름, 레이블 등)은 새 데이터 세트에 복사됩니다. 또한 파일에는 새 변수, *대치_*, 대치를 나타내는 숫자값(원래 데이터는 0, 대치된 데이터 값을 가진 케이스는 1..*n*)이 포함됩니다.

이 프로시저에서는 결과 데이터 세트가 만들어질 때 *Imputation_* 변수를 분할변수로 자동으로 정의합니다. 프로시저가 실행될 때 분할 파일이 유효하면 결과에는 분할변수 값의 각 조합에 대한 하나의 대치군이 포함됩니다.

선택적 설정

분석 가중치. 이 변수에는 분석(회귀분석 또는 표본추출) 가중치가 포함됩니다. 이 프로시저에서는 결측값 대치에 사용되는 회귀분석의 분석 가중치와 분류 모형을 통합합니다. 분석 가중치는 또한 대치된 값 요약(예: 평균, 표준 편차 및 표준 오차)에 사용됩니다. 분석 가중치가 음수이거나 0인 케이스는 제외됩니다.

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 정의된 측정 수준이 있어야 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.

방법

방법 탭은 사용된 모형 유형을 포함하여 결측값을 대치하는 방법을 지정합니다. 범주형 예측변수는 코딩된 지표(더미)입니다.

대치법. 자동 방법은 데이터를 스캔하고 데이터가 결측에 대한 단조로운 패턴을 표시할 경우 단조 방법을 사용합니다. 특정 방법을 사용할 경우 해당 방법을 **사용자 정의** 방법으로 지정할 수 있습니다.

- **전체 조건 지정 사항.** 이는 결측 데이터 패턴이 임의 패턴(단조 또는 비단조)일 경우에 사용할 수 있는 Markov chain Monte Carlo (MCMC) 방법입니다.

  변수 목록에 지정된 순서로 각 반복과 변수에 대해 전체 조건 지정(FCS) 방법은 모형에서 사용 가능한 다른 모든 변수를 예측변수로 사용하여 일변량(단일 종속변수) 모형을 적합시킨 다음 적합시키는 변수에 대한 결측값을 대치합니다. 방법은 최대 반복 수에 도달할 때까지 계속하고 최대 반복에서 대치된 값은 대치된 데이터 세트에 저장됩니다.

최대반복수. 이는 FCS 방법에 사용되는 Markov chain으로 구하는 반복 수 또는 "단계 수"를 지정합니다. FCS 방법이 자동으로 선택된 경우 10개의 반복 수를 사용합니다. FCS를 정확히 선택하면 사용자 지정 반복 수를 지정할 수 있습니다. Markov chain이 수렴되지 않은 경우 반복 수를 늘려야 합니다. 결과 탭에서 FCS 반복 히스토리 데이터를 저장하고 이를 도표로 작성하여 수렴에 액세스할 수 있습니다.

- **단조.** 이는 데이터가 단조로운 결측값 패턴을 가질 경우에만 사용할 수 있는 비반목적 방법입니다. 단조로운 패턴은 변수가 비결측값을 가질 경우 진행되는 모든 변수가 비결측값을 가지는 것처럼 변수의 순서를 지정할 수 있을 경우에 존재합니다. 이것을 **사용자 정의** 방법으로 지정할 때 단조로운 패턴을 보여주는 순서로 목록에 변수를 지정해야 합니다.

  단조로운 순서의 각 변수에 대해 단조 방법은 모형에서 모든 이전 변수를 예측변수로 사용하여 일변량(단일 종속변수) 모형을 적합시킨 다음 적합시키는 변수에 대해 결측값을 대치합니다. 이 대치된 값은 대치된 데이터 세트에 저장됩니다.

이원 상호작용이 포함됩니다. 대치법이 자동으로 선택된 경우 각 변수에 대한 대치 모형에는 예측변수에 대한 상수항 및 주효과가 포함됩니다. 특정 방법을 선택할 경우 법주형 예측변수 중 가능한 모든 이원 상호작용을 포함시킬 수 있습니다.

척도변수에 대한 모형 유형. 대치법이 자동으로 선택된 경우 선형 회귀분석은 척도변수에 대한 일변량 모형으로 사용됩니다. 특정 방법을 선택할 경우 예측 평균 일치(PMM)을 척도변수에 대한 모형으로 선택할 수 있습니다. PMM은 회귀 모형에 의해 대치된 값을 가장 가까운 관측값에 일치시키는 선형 회귀분석의 변형입니다.

로지스틱 회귀분석은 항상 범주형 변수에 대한 일변량 모형으로 사용됩니다. 모형 유형에 관계 없이 범주형 예측변수는 코딩된 지표(더미)를 사용하여 처리됩니다.

비정칙성 허용 오차. 비정칙(또는 비가역) 행렬에는 추정 알고리즘에 심각한 문제를 일으킬 수 있는 선형 종속열이 있습니다. 거의 비정칙인 행렬은 잘못된 결과를 초래할 수 있으므로 프로시저는 행렬식이 허용 오차보다 작은 행렬은 비정칙으로 취급합니다. 양수값을 지정합니다.

제한조건

제한조건 탭에서는 대치 과정에서 변수 역할과 대치된 척도변수 값을 적절하게 제한할 수 있습니다. 뿐만 아니라, 분석을 결측값의 최대 비율 미만인 변수로 제한할 수 있습니다.

변수 요약 데이터 스캔. 데이터 스캔을 클릭하면 목록은 분석 변수 및 각 분석 변수에 대한 관측 퍼센트 결측값, 최소값 및 최대값을 표시합니다. 요약은 케이스 입력란에 지정된 바에 따라 모든 케이스를 기준으로 하거나 첫 번째 *n* 케이스로 제한됩니다. **데이터 재스캔**을 클릭하면 분포 요약 정보가 업데이트됩니다.

제한조건 정의

- **역할.** 이렇게 하면 예측변수로 대치되거나 처리될 변수 세트를 사용자 정의할 수 있습니다. 일반적으로 각 분석 변수는 대치 모형에서 종속 변수와 독립 변수로 간주됩니다. **역할**은 **예측변수로만 사용**할 변수의 대치를 취소하거나 예측변수(**대치 전용**)로 사용 중인 변수를 제외하는 데 사용할 수 있습니다. 이는 범주형 변수 또는 예측변수로만 사용되는 변수로 지정할 수 있는 유일한 제한조건입니다.

- **최소값 및 최대값.** 이러한 열에서는 척도변수에 대해 허용 가능한 최소 및 최대 대치된 값을 지정할 수 있습니다. 대치된 값이 이 범위를 벗어나면 프로시저는 범위를 벗어나지 않는 값을 찾거나 다른 값을 작성하여 그 수가 최대에 도달할 때까지 사용합니다(아래의 **최대값 사용** 참조). 방법 탭의 척도변수 모형 유형으로 **선형 회귀분석**을 선택한 경우에만 이러한 열을 사용할 수 있습니다.
- **반올림.** 일부 변수는 척도로 사용될 수 있지만 자연스럽게 추가 제한되는 값을 가질 수 있습니다. 예를 들어, 한 가정 내의 사람 수는 정수가 되어야 하며 야채 상점을 방문하는 동안 지출한 금액은 분수값의 센트를 가질 수 없습니다. 이 열을 사용하면 허용할 가장 작은 액면 금액을 지정할 수 있습니다. 예를 들어, 정수 값을 얻으려면 반올림 액면 금액으로 1을 지정하고 가장 가까운 센트로 반올림할 값을 얻으려면 0.01을 지정합니다. 일반적으로 값은 가장 가까운 반올림 액면 금액의 배로 반올림됩니다. 다음 표는 여러 반올림 값이 6.64823의 대치된 값에 따라 어떻게 동작하는지 보여줍니다(반올림 전).

표 2. 반올림 결과.

반올림 액면 금액	**6.64832**가 반올림되는 값
10	10
1	7
0.25	6.75
0.1	6.6
0.01	6.65

많은 결측 데이터로 사용된 변수는 제외합니다. 일반적으로 분석 변수는 예측변수 수에 관계 없이 대치 모형을 추정하는 데 충분한 데이터를 제공하는 예측변수로 대치 및 사용됩니다. 결측값(%)이 높은 변수를 제외할 수 있습니다. 예를 들어, 50을 **최대 결측값(%)**으로 지정할 경우 50% 이상의 결측값을 가진 분석 변수는 대치되지 않고 대치 모형에서 예측변수로 사용되지 않습니다.

최대값 작성. 척도변수의 대치된 값에 최소값 및 최대값을 지정할 경우(위의 **최소값 및 최대값** 참조) 프로시저는 지정된 범위 내에 있는 일련의 값을 찾을 때까지 값을 작성합니다. 케이스별로 지정된 작성 값의 수 내에서 일련의 값을 찾지 못한 경우 프로시저는 다른 모형 모수 세트를 작성하거나 케이스 작성 프로세스를 반복합니다. 범위 내의 값을 지정된 케이스 수 및 모수 작성 수 내에서 찾지 못한 경우 오차가 발생합니다.

이 값을 증가시키면 진행 시간이 증가할 수 있습니다. 프로시저 진행 시간이 오래 걸리거나 프로시저가 적당한 작성 수를 찾지 못할 경우 지정된 최소값 및 최대갑을 검사하여 해당 값이 적절한지 확인합니다.

결과

출력. 결과 표시를 제어합니다. 전체 대치 요약은 항상 표시되며 여기에는 대치 지정 사항, (전체 조건 지정하항 방법에 대한) 반복 수, 대치된 종속변수, 대치에서 제외된 종속변수, 및 대치 시퀀스와 관련된 테이블이 포함됩니다, 지정된 경우 분석 변수에 대한 제한도 표시됩니다.

- **대치 모형.** 종속변수 및 예측변수에 대한 대치 모형을 비롯해 일변량 모형 유형, 모형 효과 및 대치된 값 수가 표시됩니다.
- **기술통계.** 대치된 변수에 대한 종속변수의 기술 통계가 표시됩니다. 척도변수의 경우 기술 통계에는 (대치 이전의) 원래 입력 데이터에 대한 평균, 개수, 표준 편차, 최소값 및 최대값, 대치된 값(대치 기준) 및 완성

데이터(대치 기준으로, 원래 값과 대치된 값 모두)가 포함됩니다. 범주형 변수의 경우 기술 통계는 원래 입력 데이터(대치 이전), 대치된 값(대치 기준), 완성 데이터(대치 기준으로 원래 값 및 대치된 값 모두)에 대한 범주별 개수 및 백분율을 포함합니다.

반복 히스토리. 전체 조건 지정 사항 대치법을 사용할 경우 FCS 대치에 대한 반복 히스토리 데이터가 포함된 데이터 세트를 요청할 수 있습니다. 데이터 세트에는 값이 대치된 각 척도 종속변수에 대한 반복 및 대치별 평균 및 표준 편차가 포함됩니다. 데이터를 작성하여 모형 수렴 액세스에 도움을 얻을 수 있습니다. 자세한 정보는 주제를 참조하십시오.

MULTIPLE IMPUTATION 명령 추가 기능

명령 구문을 사용하여 수행할 수 있는 추가 기능은 다음과 같습니다.

- 기술 통계가 표시되는 변수 서브세트를 지정합니다(IMPUTATIONSUMMARIES 하위 명령).
- 프로시저 단일 실행의 결측값 패턴 분석 및 대치를 지정합니다.
- 변수를 대치할 때 허용되는 모형 모수의 최대 수를 지정합니다(MAXMODELPARAM 키워드).

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

다중 대치된 데이터 작업

다중 대치(MI) 데이터 세트가 작성되면 변수 레이블이 *대치 번호*인 *대치_*가 추가되고 데이터 세트가 오름차순으로 정렬됩니다. 원래 데이터 세트의 케이스 값은 0입니다. 대치된 값에 대한 케이스에는 1에서 *M*까지 번호가 매겨집니다. 여기서 *M*은 대치 수입니다.

데이터 세트를 열 때 *대치_*가 존재할 경우 데이터 세트를 가능한 MI 데이터 세트로 식별합니다.

분석용 다중 대치된 데이터 세트 활성화

분석 시 데이터 세트를 MI 데이터 세트로 처리되도록 하려면 *대치_*를 가진 **그룹 비교** 옵션을 사용하여 분할해야 합니다. 또한 다른 변수의 분할을 정의할 수 있습니다.

메뉴에서 다음을 선택합니다.

데이터 > 분할 파일...

1. **그룹 비교**를 선택합니다.
2. 케이스를 그룹화할 변수로 *대치 번호[대치_]*로 선택합니다.

또는 표시를 켜면(아래 참조) 파일은 *대치 번호[대치_]*에 따라 분할됩니다.

관측값에서 대치된 값 구별하기

셀 배경 색상, 글꼴 및 굵기 유형(대치된 값에 대해)을 기준으로 관측값에서 대치된 값을 구분할 수 있습니다. 대치 결측값을 사용하여 현재 세션에 새로운 데이터 세트를 만들면 기본적으로 표시가 켜집니다. 대치가 들어 있는 저장된 데이터 파일을 열면 표시가 꺼집니다.

표시를 활성화하려면 데이터 편집기 메뉴에서 다음을 선택합니다.

보기 > 대치된 데이터 표시...

또는 데이터 편집기에서 데이터 보기에 있는 편집 막대의 오른쪽 끝에 있는 대치 표시 단추를 클릭하여 표시를 활성화할 수 있습니다.

대치 간 이동

1. 메뉴에서 다음을 선택합니다.

   편집 > 대치로 이동...
2. 드롭다운 목록에서 대치(또는 원래 데이터)를 선택합니다.

또는 데이터 편집기에서 데이터 보기의 편집 막대에 있는 드롭다운 목록에서 대치를 선택할 수 있습니다.

대치를 선택하면 상대적인 케이스 위치가 유지됩니다. 예를 들어, 원본 데이터 세트에 1000개의 케이스가 있는 경우 첫 번째 대치에 있는 케이스 1034(34번째 케이스)가 격자 맨 위에 표시됩니다. 드롭다운에서 대치 **2**를 선택한 경우 대치 2에서 34번째 케이스인 케이스 2034는 격자 맨 위에 표시됩니다. 드롭다운에서 **원래 데이터**를 선택한 경우 케이스 34는 격자 맨 위에 표시됩니다. 대치 사이를 탐색할 경우 대치 위치가 유지되어 대치 간의 값을 손쉽게 비교할 수 있습니다.

대치된 값 전송 및 편집

때때로 대치된 데이터에 대한 전송을 수행해야 합니다. 예를 들어, 급여 변수의 모든 값에 대한 로그를 가져와 새 변수에 결과를 저장합니다. 대치된 데이터를 사용하여 대치된 값을 사용하여 계산한 값이 원래 데이터를 사용하여 계산한 값과 다를 경우 대치된 데이터를 사용하여 계산한 값은 대치된 것으로 처리됩니다.

데이터 편집기의 특정 셀에 있는 대치된 값을 편집할 경우 해당 셀은 계속 대치된 것으로 처리됩니다. 이러한 방식으로 대치된 값을 편집하는 것은 권장하지 않습니다.

다중 대치된 데이터 분석

많은 프로시저에서는 다중으로 대치된 데이터 세트의 분석 결과를 풀링하는 작업을 지원합니다. 대치 표시가 활성화되어 있는 경우 풀링을 지원하는 프로시저 다음에 특별한 아이콘이 표시됩니다. 분석 메뉴의 기술 통계 하위 메뉴인 빈도, 기술, 탐색 및 교차 분석표는 모두 풀링을 지원하지만 비율, P-p 도표와 Q-Q 도표에서는 풀링을 지원하지 않습니다.

표 형식 결과와 모형 PMML은 모두 풀링될 수 있습니다. 풀링된 결과 요청에 대한 새 프로시저가 없는 대신 옵션 대화 상자에 있는 새 탭이 다중 대치 결과에 대한 광역 차원의 제어를 제공합니다.

- **표 형식 결과의 풀링.** 기본적으로 다중 대치(MI) 데이터 세트에 지원되는 프로시저를 실행할 경우 각각의 대치, 원래(대치되지 않은) 데이터 및 대치 전체의 변동을 고려하는 풀링(최종) 결과에 대한 결과가 자동으로 생성됩니다. 풀링된 통계는 절차마다 다릅니다.

- **PMML 풀링.** 또한 PMML을 내보내는 지원 프로시저에서 풀링된 PMML을 구할 수 있습니다. 풀링 PMML은 비풀링 PMML과 같은 방식으로 요청되며 비풀링 PMML 대신에 저장됩니다.

지원되지 않는 프로시저에서는 풀링 결과 또는 풀링 PMML 파일을 생성할 수 없습니다.

풀링 수준

결과는 다음과 같은 두 가지 수준 중 하나를 사용하여 풀링됩니다.

- **Naïve 조합.** 풀링된 모수만 사용할 수 있습니다.
- **일변량 조합.** 가능한 경우 풀링된 모수, *p* 값, 신뢰구간 및 풀링 진단(결측값 분수 정보, 상대적 효율성, 분산의 상대적 증가)가 표시됩니다.

계수(회귀분석 및 상관), 평균(및 평균 차이) 및 개수가 일반적으로 풀링됩니다. 통계의 표준 오차를 사용할 수 있으면 일변량 풀링이 사용되고, 그렇지 않으면 naïve 풀링이 사용됩니다.

풀링을 지원하는 절차

다음과 같은 프로시저에서는 결과의 각 부분에 대해 지정된 풀링 수준에서 MI 데이터 세트를 지원합니다.

빈도분석. 다음 기능이 지원됩니다.

- 통계 테이블은 일변량 풀링의 평균값과(표준 오차 평균도 요청된 경우) Naïve 풀링의 유효 N 및 결측 N을 지원합니다.
- 빈도 테이블은 Naïve 풀링의 개수를 지원합니다.

기술통계. 다음 기능이 지원됩니다.

- 기술통계 테이블은 일변량 풀링의 평균값과(표준 오차 평균도 요청된 경우) Naïve 풀링의 N을 지원합니다.

교차 분석표. 다음 기능이 지원됩니다.

- 교차 분석표는 Naïve 풀링의 개수를 지원합니다.

평균. 다음 기능이 지원됩니다.

- 보고서 테이블은 일변량 풀링의 평균값과(표준 오차 평균도 요청된 경우) Naïve 풀링의 N을 지원합니다.

일표본 T 검정. 다음 기능이 지원됩니다.

- 통계 테이블은 일변량 풀링의 평균값과 Naïve 풀링의 N을 지원합니다.
- 검정 테이블은 일변량 풀링의 평균차를 지원합니다.

독립 표본 T 검정. 다음 기능이 지원됩니다.

- 그룹 통계 테이블은 일변량 풀링의 평균값과 Naïve 풀링의 N을 지원합니다.
- 검정 테이블은 일변량 풀링의 평균차를 지원합니다.

대응표본 T 검정. 다음 기능이 지원됩니다.

- 통계 테이블은 일변량 풀링의 평균값과 Naïve 풀링의 N을 지원합니다.

- 상관관계표는 상관과 Naïve 풀링의 N을 지원합니다.
- 검정 테이블은 일변량 풀링의 평균값을 지원합니다.

일원 분산 분석. 다음 기능이 지원됩니다.

- 기술통계량 테이블은 일변량 풀링의 평균값과 Naïve 풀링의 N을 지원합니다.
- 대비검정 테이블은 일변량 풀링의 대비 값을 지원합니다.

선형 혼합 모형. 다음 기능이 지원됩니다.

- 기술통계 테이블은 Naïve 풀링의 평균값과 N을 지원합니다.
- 고정 효과 추정값 테이블은 일변량 풀링에서 추정값을 지원합니다.
- 공분산 모수 추정값 테이블은 일변량 풀링에서 추정값을 지원합니다.
- 추정값 주변 평균: 추정값 테이블은 일변량 풀링의 평균값을 지원합니다.
- 추정값 주변 평균: 쌍대 비교 테이블은 일변량 풀링의 평균차를 지원합니다.

일반화 선형 모형과 일반화된 추정 방정식. 이러한 절차는 풀링된 PMML을 지원합니다.

- 범주형 변수 정보 테이블은 Naïve 풀링에서 N과 퍼센트를 지원합니다.
- 연속형 변수 정보 테이블은 Naïve 풀링에서 N과 평균을 지원합니다.
- 모수 추정값 테이블은 일변량 풀링에서 계수 B를 지원합니다.
- 추정값 주변 평균: 추정계수는 Naïve 풀링에서 평균을 지원합니다.
- 추정값 주변 평균: 추정값 테이블은 일변량 풀링의 평균값을 지원합니다.
- 추정값 주변 평균: 쌍대 비교 테이블은 일변량 풀링의 평균차를 지원합니다.

이변량 상관계수. 다음 기능이 지원됩니다.

- 기술통계 테이블은 Naïve 풀링의 평균값과 N을 지원합니다.
- 상관관계표는 상관과 일변량 풀링의 N을 지원합니다. 상관은 풀링 전에 Fisher의 *z* 변환을 사용하여 변환된 다음 풀링 후 역변환됩니다.

편상관. 다음 기능이 지원됩니다.

- 기술통계 테이블은 Naïve 풀링의 평균값과 N을 지원합니다.
- 상관관계표는 Naïve 풀링의 상관을 지원합니다.

선형 회귀분석. 이 절차는 풀링된 PMML을 지원합니다.

- 기술통계 테이블은 Naïve 풀링의 평균값과 N을 지원합니다.
- 상관관계표는 상관과 Naïve 풀링의 N을 지원합니다.
- 계수 테이블은 일변량 풀링에서 B를 지원하고 Naïve 풀링에서 상관을 지원합니다.
- 상관관계표는 Naïve 풀링에서 상관을 지원합니다.
- 잔차 통계 테이블은 Naïve 풀링의 평균과 N을 지원합니다.

이분형 로지스틱 회귀분석. 이 절차는 풀링된 PMML을 지원합니다.

- 방정식의 변수 테이블은 일변량 풀링의 B를 지원합니다.

다항 로지스틱 회귀분석. 이 절차는 풀링된 PMML을 지원합니다.

- 모수 추정값 테이블은 일변량 풀링에서 계수 B를 지원합니다.

순서 회귀분석. 다음 기능이 지원됩니다.

- 모수 추정값 테이블은 일변량 풀링에서 계수 B를 지원합니다.

판별 분석. 이 절차는 풀링된 모형 XML을 지원합니다.

- 그룹 통계 테이블은 Naïve 풀링의 평균과 유효수를 지원합니다.
- 풀링 그룹-내 행렬 테이블은 Naïve 풀링에서 상관을 지원합니다.
- 정준 판별 함수 계수 테이블은 Naïve 풀링에서 비표준화 계수를 지원합니다.
- 그룹 중심값의 함수 테이블은 Naïve 풀링에서 비표준화 계수를 지원합니다.
- 분류 함수 계수 테이블은 Naïve 풀링에서 계수를 지원합니다.

카이제곱 검정. 다음 기능이 지원됩니다.

- 기술통계 테이블은 Naïve 풀링의 평균과 N을 지원합니다.
- 빈도표는 Naïve 풀링의 관측 수를 지원합니다.

이항검정. 다음 기능이 지원됩니다.

- 기술통계 테이블은 Naïve 풀링의 평균과 N을 지원합니다.
- 검정 테이블은 Naïve 풀링에서 N, 관측 비율 및 검정비율을 지원합니다.

런 검정. 다음 기능이 지원됩니다.

- 기술통계 테이블은 Naïve 풀링의 평균과 N을 지원합니다.

일표본 Kolmogorov-Smirnov 검정. 다음 기능이 지원됩니다.

- 기술통계 테이블은 Naïve 풀링의 평균과 N을 지원합니다.

독립 2-표본 비모수 검정. 다음 기능이 지원됩니다.

- 순위 테이블은 Naïve 풀링의 평균 순위와 N을 지원합니다.
- 빈도표는 Naïve 풀링의 N을 지원합니다.

독립 K-표본 비모수 검정. 다음 기능이 지원됩니다.

- 순위 테이블은 Naïve 풀링의 평균 순위와 N을 지원합니다.
- 빈도표는 Naïve 풀링의 개수를 지원합니다.

대응 2-표본 비모수 검정. 다음 기능이 지원됩니다.

- 순위 테이블은 Naïve 풀링의 평균 순위와 N을 지원합니다.

• 빈도표는 Naïve 풀링의 N을 지원합니다.

대응 K-표본 비모수 검정. 다음 기능이 지원됩니다.

• 순위 테이블은 Naïve 풀링의 평균 순위를 지원합니다.

Cox 회귀분석. 이 절차는 풀링된 PMML을 지원합니다.

• 방정식의 변수 테이블은 일변량 풀링의 B를 지원합니다.

• 공변량 평균 테이블은 Naïve 풀링에서 평균을 지원합니다.

다중 대치 옵션

다중 대치 탭은 다중 대치와 관련된 두 가지 선호도를 제어합니다.

대치된 데이터 선호도 기본적으로 대치된 데이터가 포함된 셀에는 대치되지 않은 데이터를 포함하는 셀이 될 다른 배경색이 있습니다. 대치된 데이터의 모양은 데이터 세트를 스크롤하여 해당 셀을 쉽게 찾을 수 있도록 특색이 있어야 합니다. 기본 셀 배경색, 글꼴을 변경할 수 있으며 대치된 데이터 표시를 굵은 글꼴로 만들 수 있습니다.

분석 결과 이 그룹은 여러 개로 대치된 데이터 세트를 분석할 때마다 생성되는 뷰어 결과 유형을 제어합니다. 기본적으로 축력결과는 원래(사전 대치) 데이터 세트 및 대치된 각 데이터 세트용으로 생성됩니다. 또한 대치된 데이터 풀링을 지원하는 이러한 프로시저에서 최종 풀링된 결과가 생성됩니다. 일변량 풀링이 수행될 때 풀링 진단도 표시됩니다. 그러나 보고 싶지 않은 결과를 표시하지 않을 수 있습니다.

다중 대치 옵션을 설정하려면

메뉴에서 다음을 선택합니다.

편집 > 옵션

다중 대치 탭을 클릭합니다.

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 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 라이센스까지 부여하는 것은 아닙니다. 라이센스에 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 라이센스 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 라이센스 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 라이센스가 부여된 프로그램 및 프로그램에 대해 사용 가능한 모든 라이센스가 부여된 자료는 IBM이 IBM 기본 계약, IBM 국제 프로그램 라이센스 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 검정하지 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.

저작권 라이센스:

이 정보에는 여러 운영 플랫폼에서의 프로그래밍 기법을 보여주는 원어로 된 샘플 응용프로그램이 들어 있습니다. 귀하는 이러한 샘플 프로그램의 작성 기준이 된 운영 플랫폼의 응용프로그램 프로그래밍 인터페이스(API)에 부합하는 응용프로그램을 개발, 사용, 판매 또는 배포할 목적으로 추가 비용 없이 이들 샘플 프로그램을 어떠한 형태로든 복사, 수정 및 배포할 수 있습니다. 이러한 샘플 프로그램은 모든 조건하에서 완전히 검정된 것

은 아닙니다. 따라서 IBM은 이러한 프로그램의 신뢰성, 서비스 가능성 또는 기능을 보증하거나 진술하지 않습니다. 샘플 프로그램은 어떠한 종류의 보증 없이 "현상태대로" 제공합니다. IBM은 샘플 프로그램 사용으로 인한 어떠한 손해에 대하여도 책임을 지지 않습니다.

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.

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(www.ibm.com/legal/copytrade.shtml)의 "저작권 및 상표 정보"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