

IBM SPSS Statistics

빠른 시작 안내서

주요 내용

- 데이터세트를 검사 및 정리하는 데 사용되는 데이터 검증 및 중복 사례 식별 절차를 학습합니다.
 - 온라인 도움말 메뉴에서 유용한 사례 연구를 참고합니다.
 - 더 많은 기술 개발을 위한 공개 교육 과정을 만나보십시오.
-

소개

IBM® SPSS® Statistics는 보고서 및 그래프 작성, 통계 등 일반적인 목적의 데이터 분석을 위한 통합 소프트웨어 제품군입니다.

스프레드시트에서 처리하기 어려운 대량의 데이터를 처리하고 복잡한 스프레드시트 작성에 소요되는 시간을 줄이면서 높은 품질의 결과를 얻기 위해 많은 스프레드시트 사용자들이 SPSS Statistics를 활용하고 있습니다. IBM SPSS Statistics는 통계학자들이 개발한 다양한 범위의 보고서 및 통계 분석 절차를 갖추고 있으며, 산출되는 결과의 품질을 보장하기 위해 전문가의 평가를 거쳤습니다.

이 안내서는 데이터세트를 검사 및 정리하기 위해 사용되는 *데이터 검증 및 중복 사례 식별* 절차를 소개합니다.



데이터 검증

1단계

데이터 검증 절차를 통해 확인할 수 없거나 유효하지 않은 사례, 변수 및 데이터 값을 식별할 수 있습니다.

- telco_missing.sav 파일을 엽니다. 이 파일은 C: 드라이브 또는 프로그램 설치 드라이브의 ProgramFiles/IBM/SPSS/23/Samples/English에 있습니다.
- Statistics 데이터 편집기 창의 메뉴에서 *데이터*를 클릭합니다.
- *검증*으로 이동합니다.
- *사전 정의 규칙 로드*를 클릭합니다. [참고: 이 과정은 귀하가 신규 사용자일 때 해당합니다. Statistics는 기존의 사전 정의 규칙이 대체될 수 있다는 경고 메시지를 표시합니다. 귀하에게는 사용자 지정 규칙이 없으므로 이 메시지를 무시하십시오. IBM SPSS Statistics는 표준 공통 실행 규칙의 일부를 미리 로드합니다.]
- *확인*을 클릭합니다.
- 다시 Statistics 데이터 편집기 메뉴에서 *데이터*를 클릭합니다.
- *검증*을 선택합니다.
- *데이터 검증*을 클릭합니다(그림 1).

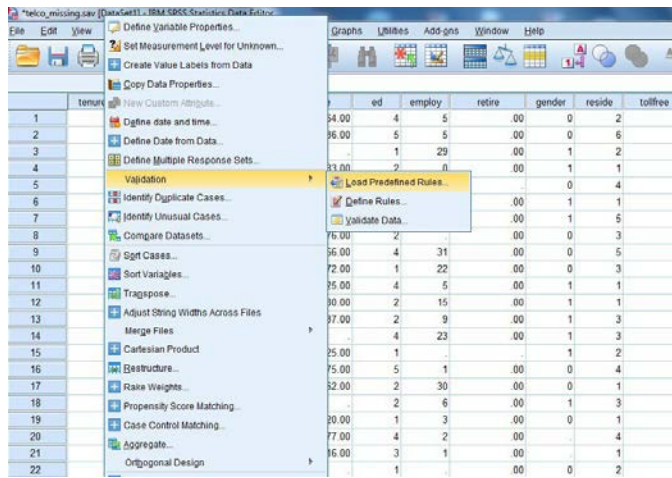


그림 1: Statistics 메뉴에서 데이터 검증을 선택합니다.

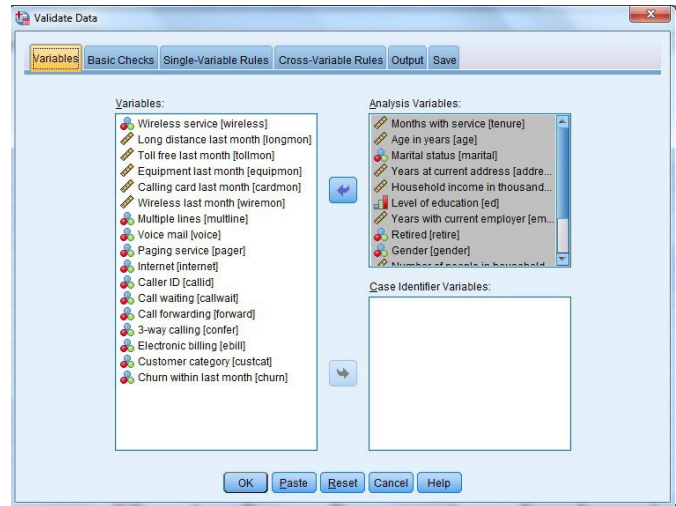


그림 2 데이터 검증 절차를 설정합니다. 데이터 검증 창에서 telco_missing.sav 파일의 변수를 *Months with service*부터 *Number of people in household*까지 선택해서 *분석 변수*로 이동시킵니다.

2단계

사용 가능한 사전 정의 규칙 모음이 표시되며, 필요에 따라 간단하거나 복잡한 기타 규칙을 추가할 수 있습니다.

- 왼쪽의 목록에서 검증하고자 하는 변수를 모두 선택합니다.
- 선택한 변수를 *분석 변수* 목록으로 이동시킵니다(그림 2).
- *단일-변수 규칙* 탭을 클릭합니다.
- 이 탭에서 분석 변수에 대한 요약 통계를 볼 수 있으며 각 변수에 기반하여 검증 규칙을 지정할 수 있습니다.
- *규칙 정의를* 클릭하여 규칙을 수정하거나 새로운 규칙을 추가할 수 있습니다. 이 탭은 살펴보기만 하고 규칙을 설정하지 마십시오.
- *계속*을 클릭합니다.

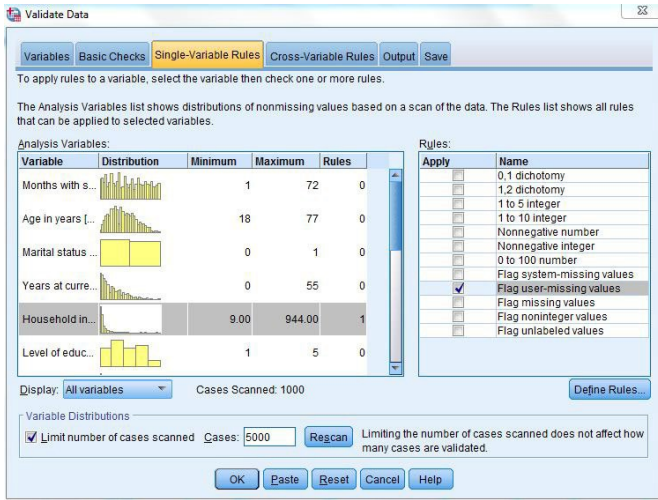


그림 3. 선택된 변수에 규칙을 적용합니다. 데이터 검증 대화 상자의 단일-변수 규칙 탭에서 분석 변수 목록에 있는 *Household income in thousands* 변수를 선택합니다. 규칙 목록에서 *사용자 결측값 플래그 지정*을 선택합니다.

2단계 (계속)

분석 변수 목록에서 변수를 한 가지 선택합니다.

- 규칙 목록에서 선택된 변수에 적용할 규칙의 상자를 선택합니다(그림 3).
- **교차-변수 규칙** 탭을 클릭합니다. [참고: 이 탭은 하나 이상의 변수에 관련된 규칙을 정의하는 데 사용됩니다. 개별적으로는 유효하지만 두 개 이상 결합되면 무효가 되는 변수 값을 논리식에서 정의합니다. 대표적인 예로, 성별: 남성과 임신 테스트 결과: 양성을 들 수 있습니다. 지금은 이 탭을 살펴보기만 하고 규칙을 설정하지 않습니다.]
- **규칙 정의**를 클릭합니다.
- **계속**을 클릭합니다.
- **확인**을 클릭합니다.

3단계

데이터 검증 절차에서는 다양한 방식으로 규칙 위반을 요약한 몇 가지 표가 생성됩니다. 사용되는 규칙을 위반하는 데이터가 없으면 아무 규칙도 위반되지 않았다는 내용의 메시지가 표시됩니다.

- **변수 요약**은 분석 변수별로 정리됩니다.
- 각 변수에는 위반된 규칙의 목록이 있습니다.
- 마지막 열에는 특정 변수에 대한 규칙 위반 횟수가 표시됩니다(그림 4).

Variable Summary

	Rule	Number of Violations
Household income in thousands	Flag missing values	179
	Total	179

Case Report^b

Case	Validation Rule Violations
	Single-Variable ^a
3	Flag missing values (1)
14	Flag missing values (1)
18	Flag missing values (1)
22	Flag missing values (1)
26	Flag missing values (1)
32	Flag missing values (1)
40	Flag missing values (1)
51	Flag missing values (1)
58	Flag missing values (1)
59	Flag missing values (1)

그림 4: 변수 요약 및 사례 보고서: 데이터 검증 대화 상자의 규칙으로 확인된 각 사례 및 요약이 결과 창에 표시됩니다.

중복 사례 식별

1단계

중복 사례 식별 절차를 통해 선택된 변수 결합에 기반하여 중복 값을 가진 사례를 확인할 수 있습니다.

- stroke_invalid.sav 파일을 엽니다. 이 파일은 기본 SPSS Statistics 설치 디렉터리에서 찾을 수 있습니다.
- Statistics 데이터 편집기 창의 메뉴에서 *데이터*를 선택합니다.
- **중복 사례 식별**을 클릭합니다(그림 5).

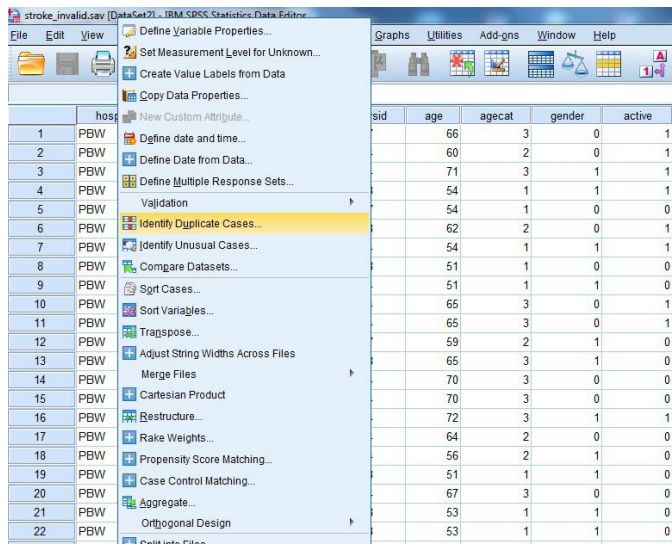


그림 5. Statistics 메뉴에서 **중복 사례 식별**을 선택합니다.

2단계

여기서 기본 사례 자동 식별에 대한 규칙을 지정할 수 있습니다. 예를 들어, 생성 날짜가 가장 최근인 사례를 기본으로 식별하고자 할 수 있습니다. 이 때 동일 집단의 이전 사례가 중복됩니다.

- 왼쪽의 목록에서 중복 사례를 정의하는 변수를 선택합니다.
- 모든 필드에서 일치하는 완전 중복 사례를 찾으려면 변수 전체를 선택합니다. 대부분의 경우, 중복 사례는 해당 데이터세트의 부분 집합만을 사용하여 정의됩니다.
- 선택한 변수를 *일치 사례 정의 기준* 목록으로 이동시킵니다(그림 6).
- **확인**을 클릭합니다.

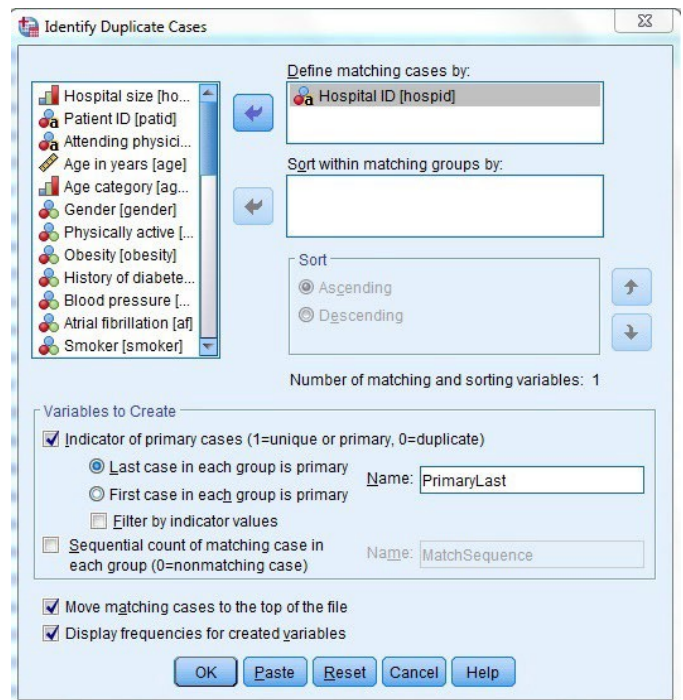


그림 6. 중복 사례 식별 절차를 설정합니다. **중복 사례 식별** 대화 상자에서 stroke_invalid.sav 파일의 몇 가지 변수를 선택하여 **일치 사례 정의 기준**으로 이동시킵니다.

3단계

데이터세트에 있는 중복 사례를 볼 수 있습니다.

- **결과** 창에서 **기본 사례인 마지막 일치 사례의 지표** 표로 결과가 요약됩니다(그림 7).
- 이는 기본 사례의 수와 중복 사례의 수를 나타내는 빈도 표입니다.
- 기본 사례는 값이 일치하는 사례 집단의 마지막 사례로 기본 정의됩니다. 이 정의는 첫 번째 사례를 대신 사용하도록 변경될 수 있으며, 다른 변수를 사용해 집단을 구분하여 다양하게 기본 사례를 정의할 수도 있습니다.

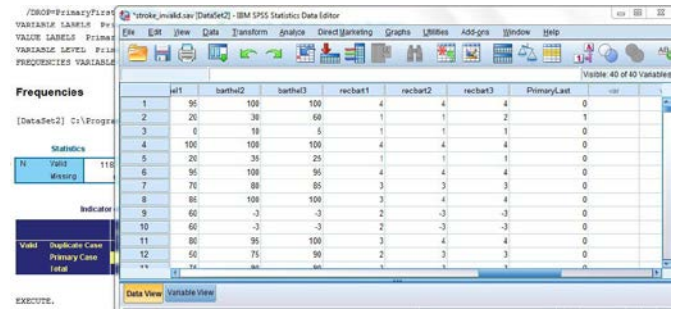


그림 8 데이터 편집기에서 확인된 중복 사례

4단계

지표 변수는 추가적인 데이터 처리에 유용하게 사용됩니다.

- 데이터세트에 사례가 기본인지 혹은 중복인지를 표시하는 새로운 변수가 생성됩니다(그림 8). [이 변수는 데이터세트에 저장할 수 있습니다. 대부분 이 지표 변수는 분석을 위해 데이터를 정렬하거나 중복 사례를 삭제하는 데 사용됩니다.]

IBM SPSS Statistics 사례 연구

데이터 검증

IBM SPSS Statistics 메뉴에서 **도움말...사례 연구**를 클릭합니다. **사례 연구**의 개요 창이 열립니다. **데이터 준비 옵션**을 확장하고 **데이터 검증**을 선택합니다.

중복 사례 식별

IBM SPSS Statistics 메뉴에서 **도움말...사례 연구**를 클릭합니다. **사례 연구**의 개요 창이 열립니다. **데이터 준비 옵션**을 확장하고 **데이터 검증**을 선택한 다음 **의료 데이터베이스 검증**을 클릭합니다.

Statistics

N	Valid	1183
	Missing	0

Indicator of each last matching case as Primary

	Frequency	Percent	Valid Percent	Cumulative Percent
Valid Duplicate Case	1167	98.6	98.6	98.6
Valid Primary Case	16	1.4	1.4	100.0
Total	1183	100.0	100.0	

그림 7 기본 사례인 마지막 일치 사례의 지표입니다. **결과** 창은 **중복 사례 식별** 대화 상자에서 일치 사례 정의에 따라 식별된 기본 및 중복 사례의 요약을 보여줍니다.

IBM Business Analytics 소개

IBM Business Analytics 소프트웨어는 앞서가는 조직의 스마트한 업무를 위해 데이터에 기반한 인사이트를 제공합니다.

비즈니스 인텔리전스, 예측 분석과 결정 관리, 성과 및 위험 관리 등을 종합한 Business Analytics 솔루션을 통해 기업들은 소비자 분석처럼 비즈니스 성과에 상당한 도움을 주는 지역 내 추세나 패턴을 발견하고 시각화할 수 있습니다.

또한, 시나리오를 비교하면서 잠재적인 위험과 기회, 더 나은 계획과 필요한 예산 및 자원을 예측할 수 있으며 기대하는 이익 대비 위험 요소를 계산하고, 규제 사항에 맞춰 업무를 수행할 수 있습니다. 조직은 광범위한 분석을 통해 경영 목표를 달성하기 위한 면밀하고 전략적인 결정을 내릴 수 있습니다. 자세한 정보를 확인하려면 ibm.com/business-analytics를 방문하십시오..

전화 상담 요청

전화 상담을 원하시거나 문의 사항이 있는 경우 [여기를 클릭](#) 하거나 02-3781-7335번으로 전화해 주십시오. IBM 상담원이 근무일 2일 내에 귀하의 문의 사항에 답변해드립니다.



© Copyright IBM Corporation 2015

IBM Corporation
IBM Analytics
Route 100
Somers, NY 10589

Produced in the United States of America
May 2015

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

이 문서는 최초 발행일을 기준으로 하며, 통지 없이 언제든지 변경될 수 있습니다. IBM이 영업하는 모든 국가에서 모든 오퍼링이 제공되는 것은 아닙니다.

이 문서의 정보는 상품성, 특정 목적에의 적합성에 대한 보증 및 타인의 권리 침해에 대한 보증이나 조건을 포함하여(단, 이에 한하지 않음) 명시적이든 묵시적이든 일체의 보증 없이 "현상태대로" 제공됩니다. IBM 제품에 대한 보증은 제품의 준거 계약 조항에 의거하여 제공됩니다.



Please Recycle