

***IBM SPSS Statistics 26***  
요약 안내서

**IBM**

**참고**

이 정보와 이 정보가 지원하는 제품을 사용하기 전에, 103 페이지의 『주의사항』에 있는 정보를 확인하십시오.

**제품 정보**

이 개정판은 새 개정판에서 별도로 명시하지 않는 한, IBM SPSS Statistics의 버전 26, 릴리스 0. 수정사항 0 및 모든 후속릴리스와 수정에 적용됩니다.

# 목차

제 1 장 소개 . . . . .	1	레이어 작성 및 표시 . . . . .	45
표본 파일 . . . . .	1	표 편집 . . . . .	45
데이터 파일 열기 . . . . .	1	행과 열 감추기 . . . . .	46
분석 실행 . . . . .	3	데이터 표시 형식 바꾸기 . . . . .	47
도표 작성 . . . . .	5	표모양 . . . . .	49
제 2 장 데이터 읽기 . . . . .	9	사전 정의된 형식 사용 . . . . .	49
데이터 파일의 IBM SPSS Statistics 기본 구조 . . . . .	9	표모양 스타일 사용자 정의 . . . . .	49
IBM SPSS Statistics 데이터 파일 읽기 . . . . .	9	기본값 표 형식 바꾸기 . . . . .	52
Excel 데이터 읽기 . . . . .	10	초기 출력 설정 사용자 정의 . . . . .	53
데이터베이스에서 데이터 읽기 . . . . .	13	변수와 값 레이블 표시 . . . . .	54
텍스트 파일에서 데이터 읽기 . . . . .	17	다른 애플리케이션에서 결과 사용 . . . . .	56
제 3 장 데이터 편집기 사용 . . . . .	21	Word 표로 결과 붙여넣기 . . . . .	56
숫자 데이터 입력 . . . . .	21	텍스트로 결과 붙여넣기 . . . . .	56
문자열 데이터 입력 . . . . .	23	Microsoft Word, PowerPoint 및 Excel 파일	
데이터 정의 . . . . .	23	로 결과 내보내기 . . . . .	57
변수 레이블 추가 . . . . .	23	PDF로 결과 내보내기 . . . . .	64
변수 유형 및 형식 바꾸기 . . . . .	24	HTML로 결과 내보내기 . . . . .	65
값 레이블 추가 . . . . .	25	제 7 장 구문 사용 . . . . .	67
결측 데이터 처리 . . . . .	26	구문 붙여넣기 . . . . .	67
숫자변수에 대한 결측값 . . . . .	27	구문 편집 . . . . .	68
문자열 변수에 대한 결측값 . . . . .	27	구문 파일 열기 및 실행 . . . . .	69
제 4 장 개별 변수의 요약 통계 탐색 . . . . .	29	중단점 사용하기 . . . . .	70
측정 수준 . . . . .	29	제 8 장 데이터 값 수정 . . . . .	73
범주형 데이터에 대한 요약 측도 . . . . .	29	척도변수에서 범주형 변수 작성 . . . . .	73
범주형 데이터에 대한 도표 . . . . .	30	새로운 변수 계산 . . . . .	75
척도변수에 대한 요약 측도 . . . . .	31	표현식에 함수 사용 . . . . .	77
척도변수에 대한 히스토그램 . . . . .	32	조건식 사용 . . . . .	77
제 5 장 도표 작성 및 편집 . . . . .	35	날짜 및 시간을 사용한 작업 . . . . .	78
도표 작성 기본 사항 . . . . .	35	두 날짜 사이의 시간 길이 계산 . . . . .	79
도표 작성기 갤러리 사용 . . . . .	36	날짜에 기간 더하기 . . . . .	80
변수 및 통계 정의 . . . . .	37	제 9 장 데이터 정렬 및 선택 . . . . .	83
텍스트 추가 . . . . .	39	데이터 정렬 . . . . .	83
도표 작성 . . . . .	39	분할 파일 처리 . . . . .	84
제 6 장 결과 작업 . . . . .	41	분할 파일 처리의 케이스 정렬 . . . . .	85
뷰어 사용 . . . . .	41	분할 파일 처리 켜기 및 끄기 . . . . .	86
피벗 표 편집기 사용 . . . . .	42	케이스 서브세트 선택 . . . . .	86
결과 정의 액세스 . . . . .	42	조건식을 기준으로 케이스 선택 . . . . .	87
피벗 표 . . . . .	43	케이스의 무작위 표본 선택 . . . . .	88
		시간 범위 또는 케이스 범위 선택 . . . . .	89
		선택하지 않은 케이스 처리 . . . . .	90

케이스 선택 상태 . . . . .	90	색인 . . . . .	107
제 10 장 표본 파일 . . . . .	91		
주의사항. . . . .	103		
상표 . . . . .	105		

---

## 제 1 장 소개

이 안내서는 제공하는 여러 기능의 사용 방법을 설명하기 위해 예제를 통한 단계별 실습을 제공합니다. 이 예제에 표시된 모든 파일은 필요한 경우 스스로 연습해 볼 수 있도록 안내서와 함께 설치되므로 동일한 분석을 수행했을 때 여기에 표시된 것과 동일한 결과를 얻을 수 있습니다.

다양한 통계 분석 기법의 자세한 예제를 보려면 도움말 메뉴에서 사용 가능한 단계별 사례 연구를 수행해 보십시오.

---

### 표본 파일

여기에 표시된 대부분의 예제에서는 *demo.sav* 데이터 파일을 사용합니다. 이 데이터 파일은 수천 명을 대상으로 한 가상 설문조사로 기본적인 인구 통계학적 정보와 소비자에 대한 정보를 포함합니다.

Student 버전을 사용 중인 경우, 이 버전에서는 *demo.sav*가 제공되는 데이터 파일의 대표적인 표본이며 이 파일의 케이스 수는 1,500개로 제한되어 있습니다. 따라서 이 데이터 파일을 사용하여 얻은 결과는 여기에 표시된 결과와 다릅니다.

제품에 설치된 표본 파일은 설치 디렉토리 표본 추출 하위 디렉토리에서 찾으실 수 있습니다. 영어, 프랑스어, 독일어, 이탈리아어, 일본어, 한국어, 폴란드어, 러시아어, 중국어 간체, 스페인어 및 중국어 번체 등의 각 언어에 대해 Samples 하위 디렉터리 내에 별도의 폴더가 있습니다.

모든 언어에서 모든 표본 파일을 사용할 수 있는 것은 아닙니다. 특정 언어로 표본 파일을 사용할 수 없는 경우 해당 언어 폴더에 영어 버전의 표본 파일이 들어 있습니다.

---

### 데이터 파일 열기

데이터 파일 열기:

1. 메뉴에서 다음을 선택합니다.

파일 > 열기 > 데이터...

파일 열기 대화 상자가 표시됩니다.

기본적으로 IBM® SPSS® Statistics 데이터 파일(확장: *.sav*)이 표시됩니다.

이 예제에서는 파일 *demo.sav*를 사용합니다.

demo.sav - Data Editor

File Edit View Data Transform Analyze Graphs Utilities Add-ons Window Help

20 : age 40

	age	marital	address	income	inccat	car
1	55	Marital status	12	72.00	3.00	36.
2	56	0	29	153.00	4.00	76.
3	28	1	9	28.00	2.00	13.
4	24	1	4	26.00	2.00	12.
5	25	0	2	23.00	1.00	11.
6	45	1	9	76.00	4.00	37.
7	42	0	19	40.00	2.00	19.
8	35	0	15	57.00	3.00	28.
9	46	0	26	24.00	1.00	12.
10	34	1	0	89.00	4.00	46.
11	55	1	17	72.00	3.00	35.

Data View Variable View

그림 1. 데이터 편집기의 *demo.sav* 파일

데이터 파일이 데이터 편집기에 표시됩니다. 데이터 보기에서 마우스 커서를 변수 이름(열 머리말) 위에 두면 해당 변수에 정의된 레이블이 있는 경우 변수 설명 레이블이 나타납니다.

기본적으로 실제 데이터 값이 표시됩니다. 레이블 표시:

2. 메뉴에서 다음을 선택합니다.

보기 > 값 레이블



그림 2. 값 레이블 단추

또는 도구 모음의 값 레이블 단추를 사용할 수도 있습니다.

	age	marital	address	income	inccat	car
1	55	Married	12	72.00	\$50 - \$74	36.
2	56	Unmarried	29	153.00	\$75+	76.
3	28	Married	9	28.00	\$25 - \$49	13.
4	24	Married	4	26.00	\$25 - \$49	12.
5	25	Unmarried	2	23.00	Under \$25	11.
6	45	Married	9	76.00	\$75+	37.
7	42	Unmarried	19	40.00	\$25 - \$49	19.
8	35	Unmarried	15	57.00	\$50 - \$74	28.
9	46	Unmarried	26	24.00	Under \$25	12.
10	34	Married	0	89.00	\$75+	46.
11	55	Married	17	72.00	\$50 - \$74	35.

그림 3. 데이터 편집기에 표시된 값 레이블

이제 구체적인 값 레이블이 표시되어 반응을 쉽게 이해할 수 있습니다.

## 분석 실행

추가 기능 옵션을 사용하는 경우, 분석 메뉴에 보고 및 통계 분석 범주의 목록이 들어 있습니다.

빈도표(개수표) 작성으로 시작합니다. 이 예를 사용하려면 Statistics Base Edition이 필요합니다.

1. 메뉴에서 다음을 선택합니다.

분석 > 기술통계량 > 빈도...

빈도분석 대화 상자가 표시됩니다.

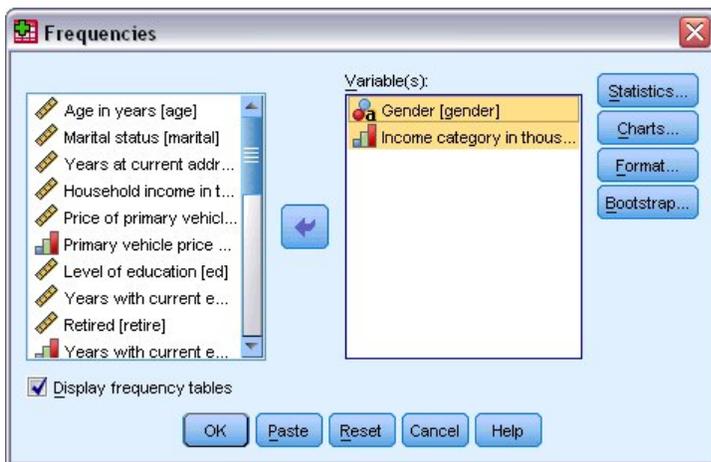


그림 4. 빈도분석 대화 상자

각 변수 옆의 아이콘은 데이터 유형 및 측정 수준에 대한 정보를 제공합니다.

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

변수 레이블 및/또는 이름이 목록에서 잘려 나타날 경우 커서를 변수 위에 두면 구체적인 레이블/이름이 표시됩니다. 변수 이름 소득범주가 대괄호 안에 변수 설명 레이블 뒤에 표시됩니다. 가구소득범주 [단위: 1000]는 변수 레이블입니다. 변수 레이블이 없을 경우 변수 이름만 목록 상자에 나타납니다.

외부 경계나 모서리를 클릭하고 끌어 창과 마찬가지로 대화 상자 크기를 조정할 수 있습니다. 예를 들어, 대화 상자를 더 넓게 만들면 변수 목록도 넓어집니다.

대화 상자 왼쪽의 소스 목록에서 분석하려는 변수를 선택하고 오른쪽의 변수 목록으로 끌어 놓습니다. **확인** 단추는 하나 이상의 변수를 변수 목록으로 옮기면 활성화되고, 이 단추를 클릭하면 분석이 실행됩니다.

많은 대화 상자에서, 목록에 있는 변수 이름을 마우스 오른쪽 단추로 클릭하고 팝업 메뉴에서 **변수 정보**를 선택하면 추가 정보를 볼 수 있습니다.

2. 소스 변수 목록에서 **성별 [성별]**을 클릭하고 변수를 대상 변수 목록으로 끌어다 놓습니다.
3. 소스 목록의 **가구소득범주 [단위: 1000] [소득범주]**를 클릭한 다음 대상 목록으로 끌어다 놓습니다.

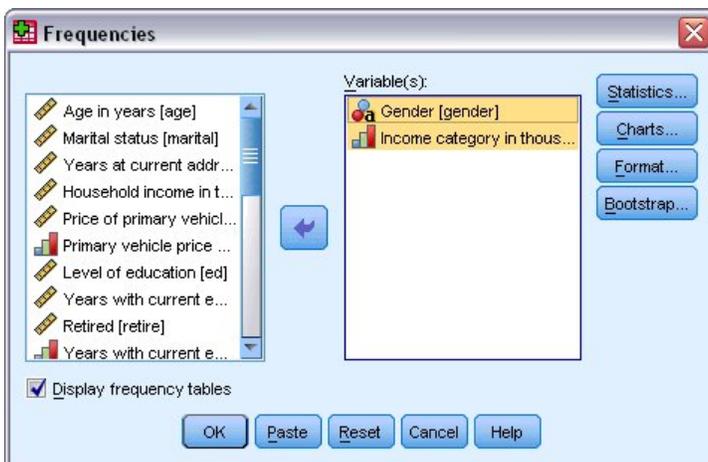


그림 5. 분석을 위해 선택한 변수

4. 확인을 클릭하여 프로시저를 실행합니다.

결과가 뷰어 창에 표시됩니다.

The screenshot shows the SPSS Output2 - Viewer window. The left pane shows a tree view with 'Income category in thousands' selected. The main area displays two tables. The first table is titled 'Gender' and the second is titled 'Income category in thousands'. Both tables show Frequency, Percent, Valid Percent, and Cumulative Percent for their respective categories.

		Frequency	Percent	Valid Percent	Cumul: Percent
Valid	Female	3179	49.7	49.7	
	Male	3221	50.3	50.3	
	Total	6400	100.0	100.0	

		Frequency	Percent	Valid Percent	Cum: Percent
Valid	Under \$25	1174	18.3	18.3	
	\$25 - \$49	2388	37.3	37.3	
	\$50 - \$74	1120	17.5	17.5	
	\$75+	1718	26.8	26.8	
	Total	6400	100.0	100.0	

그림 6. 가구소득 범주에 대한 빈도표

## 도표 작성

몇 가지 통계 프로시저를 통해 도표를 작성할 수 있지만 그래프 메뉴를 사용하여 도표를 작성할 수도 있습니다.

예를 들어, 핸드폰 서비스와 PDA(Personal Digital Assistant) 소유권 간의 관계를 보여주는 도표를 작성할 수 있습니다.

1. 메뉴에서 다음을 선택합니다.

그래프 > 도표 작성기...



그림 7. 완성된 끌어놓기 영역이 표시된 도표 작성기 대화 상자

2. 갤러리 탭을 클릭합니다(선택되어 있지 않은 경우).
3. 막대 탭을 클릭합니다(선택되어 있지 않은 경우).
4. 갤러리 위의 넓은 영역인 캔버스로 수평배열 막대도표 아이콘을 끕니다.
5. 변수 목록에서 아래로 스크롤하여 핸드폰 [핸드폰]을 마우스 오른쪽 단추로 클릭한 다음 명목을 해당 측정 수준으로 선택합니다.
6. 핸드폰 [핸드폰] 변수를  $x$ 축으로 끕니다.
7. PDA 소유 [PDA 보유]를 마우스 오른쪽 단추로 클릭한 다음 명목을 해당 측정 수준으로 선택합니다.
8. PDA 소유 [PDA 보유] 변수를 캔버스 오른쪽 위에 있는 군집 끌어놓기 영역으로 끕니다.
9. 확인을 클릭하여 도표를 작성합니다.

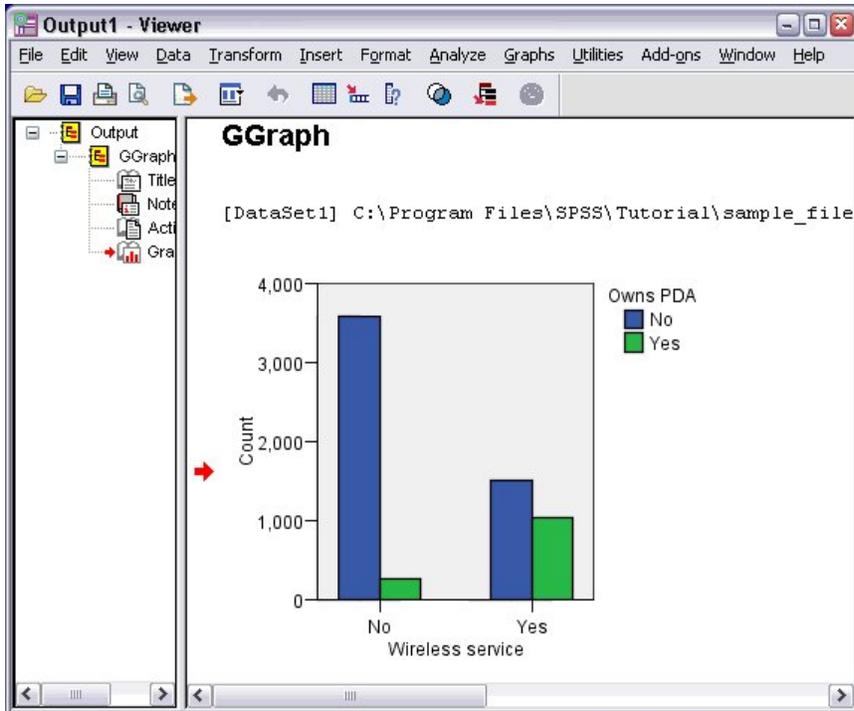


그림 8. 뷰어 창에 표시된 막대도표

막대도표가 뷰어에 표시됩니다. 막대도표는 무선 전화 서비스를 사용하는 사람이 핸드폰을 사용하지 않는 사람보다 훨씬 더 많이 PDA를 갖고 있음을 보여 줍니다.

뷰어 창의 목차 분할창에서 도표와 표를 두 번 클릭하여 편집하고 결과를 복사하여 다른 애플리케이션에 붙여넣을 수 있습니다. 이러한 항목은 뒷부분에서 다룹니다.



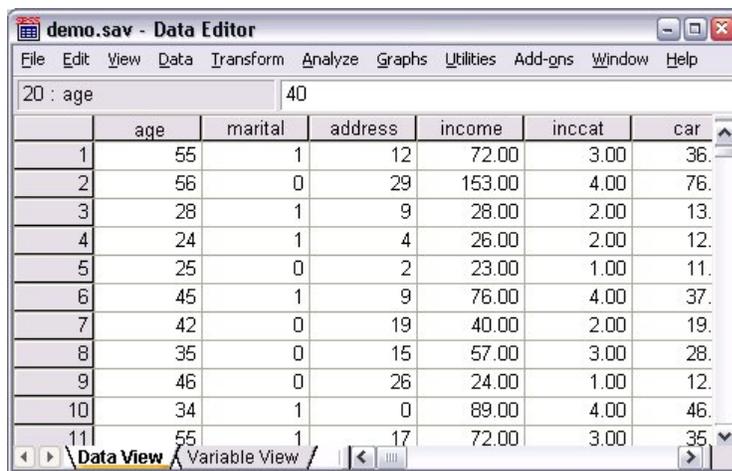
---

## 제 2 장 데이터 읽기

데이터는 직접 입력하거나 여러 다른 소스에서 가져올 수 있습니다. 이 장에서는 IBM SPSS Statistics 데이터 파일, Microsoft Excel과 같은 스프레드시트 애플리케이션, Microsoft Access와 같은 데이터베이스 애플리케이션 및 텍스트 파일에 저장된 데이터 읽기 프로세스를 설명합니다.

---

### 데이터 파일의 IBM SPSS Statistics 기본 구조



	age	marital	address	income	inccat	car
1	55	1	12	72.00	3.00	36.
2	56	0	29	153.00	4.00	76.
3	28	1	9	28.00	2.00	13.
4	24	1	4	26.00	2.00	12.
5	25	0	2	23.00	1.00	11.
6	45	1	9	76.00	4.00	37.
7	42	0	19	40.00	2.00	19.
8	35	0	15	57.00	3.00	28.
9	46	0	26	24.00	1.00	12.
10	34	1	0	89.00	4.00	46.
11	55	1	17	72.00	3.00	35.

그림 9. 데이터 편집기

IBM SPSS Statistics 데이터 파일은 케이스(행)와 변수(열)로 구성됩니다. 이 데이터 파일에서 케이스는 설문조사에 반응한 개별 응답자를 나타냅니다. 변수는 설문조사에 있는 각 질문 사항에 대한 반응을 나타냅니다.

---

### IBM SPSS Statistics 데이터 파일 읽기

IBM SPSS Statistics 데이터 파일은 파일 확장이 *.sav*이며 저장된 데이터를 포함합니다.

1. 메뉴에서 다음을 선택합니다.

파일 > 열기 > 데이터...

2. *demo.sav*를 찾은 다음 이 파일을 엽니다. 자세한 정보는 91 페이지의 제 10 장 『표본 파일』 주제를 참조하십시오.

데이터 편집기에 데이터가 표시됩니다.

	age	marital	address	income	inccat	car
1	55	1	12	72.00	3.00	36.
2	56	0	29	153.00	4.00	76.
3	28	1	9	28.00	2.00	13.
4	24	1	4	26.00	2.00	12.
5	25	0	2	23.00	1.00	11.
6	45	1	9	76.00	4.00	37.
7	42	0	19	40.00	2.00	19.
8	35	0	15	57.00	3.00	28.
9	46	0	26	24.00	1.00	12.
10	34	1	0	89.00	4.00	46.
11	55	1	17	72.00	3.00	35.

그림 10. 열려 있는 데이터 파일

## Excel 데이터 읽기

데이터 편집기에 모든 데이터를 직접 입력하지 않고 Microsoft Excel과 같은 애플리케이션에서 데이터를 읽을 수 있습니다. 열 머리말을 변수 이름으로 읽을 수도 있습니다.

1. 메뉴에서 다음을 선택합니다.

파일 > 데이터 가져오기 > Excel

2. Samples\English 폴더로 이동하여 demo.xlsx를 선택합니다.

Excel 파일 읽기 대화 상자에 데이터 파일의 미리보기가 표시됩니다. 파일에 있는 첫 번째 시트의 콘텐츠가 표시됩니다. 파일에 여러 개의 시트가 있는 경우 목록에서 시트를 선택할 수 있습니다.

*Gender*의 문자열 값 중 일부에는 선행 공백이 있습니다. *MaritalStatus*의 값 중 일부는 마침표(.)로 표시됩니다.

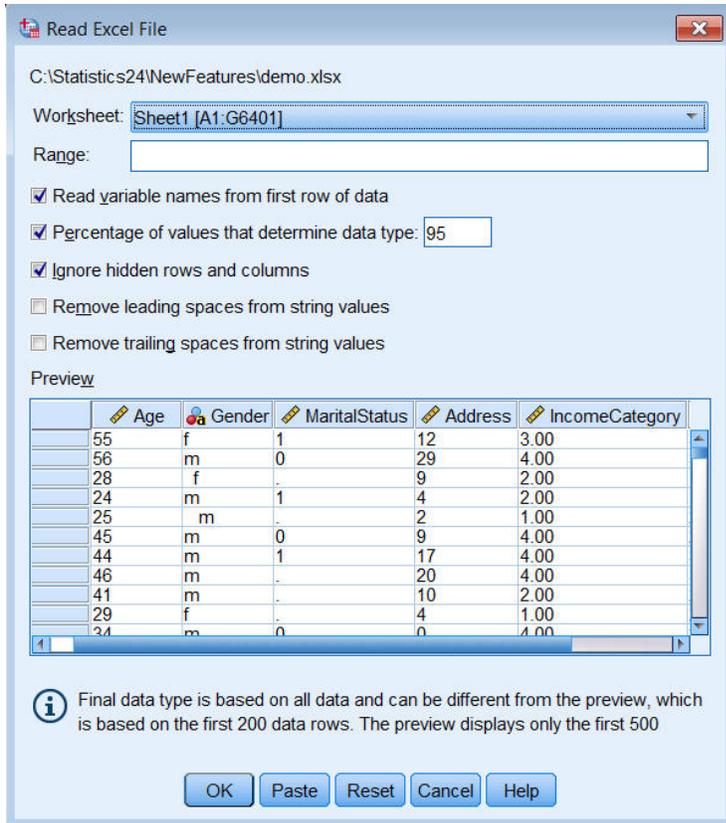
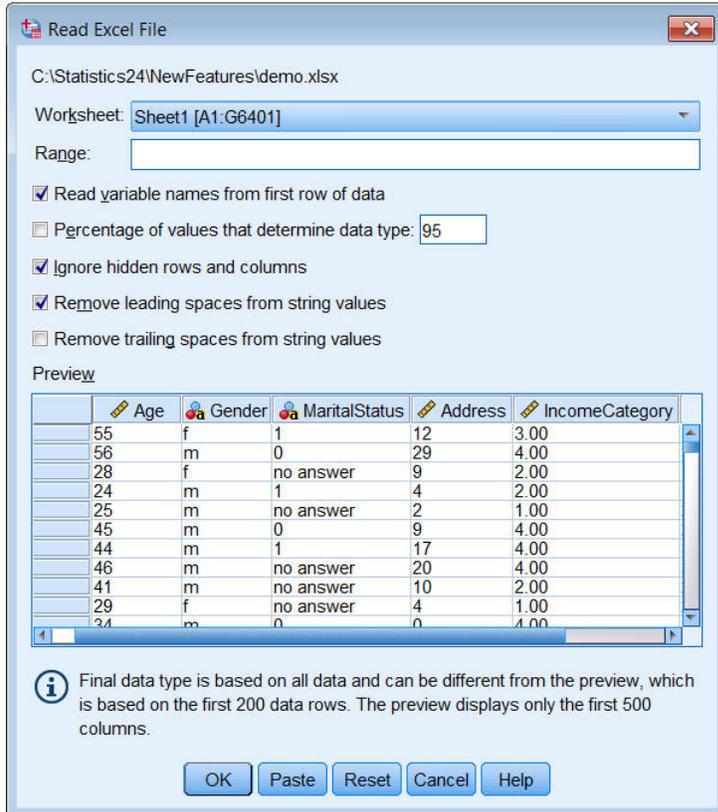


그림 11. Excel 파일 읽기 대화 상자

3. 데이터 첫 행에서 변수 이름 읽어오기가 선택되었는지 확인합니다. 열 머리가 변수 이름 규칙을 따르지 않는 경우 열 머리는 유효한 변수 이름으로 변환됩니다. 원래 열 머리는 변수 레이블로 저장됩니다.
4. 문자열 값에서 선행 공백 제거를 선택합니다.
5. 데이터 유형을 결정하는 값의 퍼센트를 선택 취소합니다.



시스템 결측값인 셀에 이제 문자열 값 "응답 없음"이 표시됩니다. 값 퍼센트 매개변수가 있고 열에 혼합 데이터 유형이 포함되는 경우 변수는 문자열 데이터 유형으로 읽혀집니다. 모든 값이 보존되지만 숫자값은 문자열 값으로 처리됩니다.

6. 데이터 유형을 결정하는 값의 퍼센트를 선택하여 *MaritalStatus*를 숫자 변수로 처리합니다.
7. 확인을 클릭하여 Excel 파일을 읽습니다.

이제 열 머리말이 변수 이름으로 사용된 데이터가 데이터 편집기에 나타납니다. 변수 이름은 공백을 포함할 수 없으므로 원래 열 머리말에서 공백이 제거됩니다. 예를 들어, 열 머리말 "결혼 상태"가 *MaritalStatus* 변수로 변환됩니다. 원래 열 머리말은 변수 레이블로 유지됩니다.

The screenshot shows the IBM SPSS Statistics Data Editor interface. The title bar reads '\*Untitled2 [DataSet1] - IBM SPSS Statistics Data Editor'. The menu bar includes File, Edit, View, Data, Transform, Analyze, Direct Marketing, Graphs, Utilities, Extensions, Window, and Help. The toolbar contains various icons for file operations and data manipulation. The main window displays a data table with 10 rows and 6 columns. The columns are labeled Age, Gender, MaritalStatus, Address, IncomeCategory, and JobCategory. The data is as follows:

	Age	Gender	MaritalStatus	Address	IncomeCategory	JobCategory
1	55	f	1	12	3.00	3
2	56	m	0	29	4.00	3
3	28	f	.	9	2.00	1
4	24	m	1	4	2.00	1
5	25	m	.	2	1.00	2
6	45	m	0	9	4.00	2
7	44	m	1	17	4.00	3
8	46	m	.	20	4.00	3
9	41	m	.	10	2.00	2
10	29	f	.	4	1.00	2

At the bottom of the window, there are tabs for 'Data View' (selected) and 'Variable View'.

그림 12. 가져온 Excel 데이터

**관련 정보:**

91 페이지의 제 10 장 『표본 파일』

## 데이터베이스에서 데이터 읽기

데이터베이스 마법사를 사용하여 데이터베이스 소스에서 데이터를 쉽게 가져올 수 있습니다. ODBC(Open Database Connectivity) 드라이버를 설치하면 이 드라이버를 사용하는 데이터베이스를 직접 읽을 수 있습니다. 여러 데이터베이스 형식의 ODBC 드라이버가 설치 CD에 제공됩니다. 타공급 업체의 드라이버를 추가할 수도 있습니다. 가장 많이 사용되는 데이터베이스 애플리케이션 중 하나인 Microsoft Access를 이 예제에서 설명합니다.

참고: 이 예제는 Microsoft Windows에 관련한 예제이며 Access용 ODBC 드라이버가 필요합니다. Microsoft Access ODBC 드라이버는 32비트 버전의 IBM SPSS Statistics에서만 작동합니다. 단계는 다른 플랫폼에서와 유사하지만 타사의 Access용 ODBC 드라이버가 필요할 수 있습니다.

1. 메뉴에서 다음을 선택합니다.

파일 > 데이터 가져오기 > 데이터베이스 > 새 쿼리...

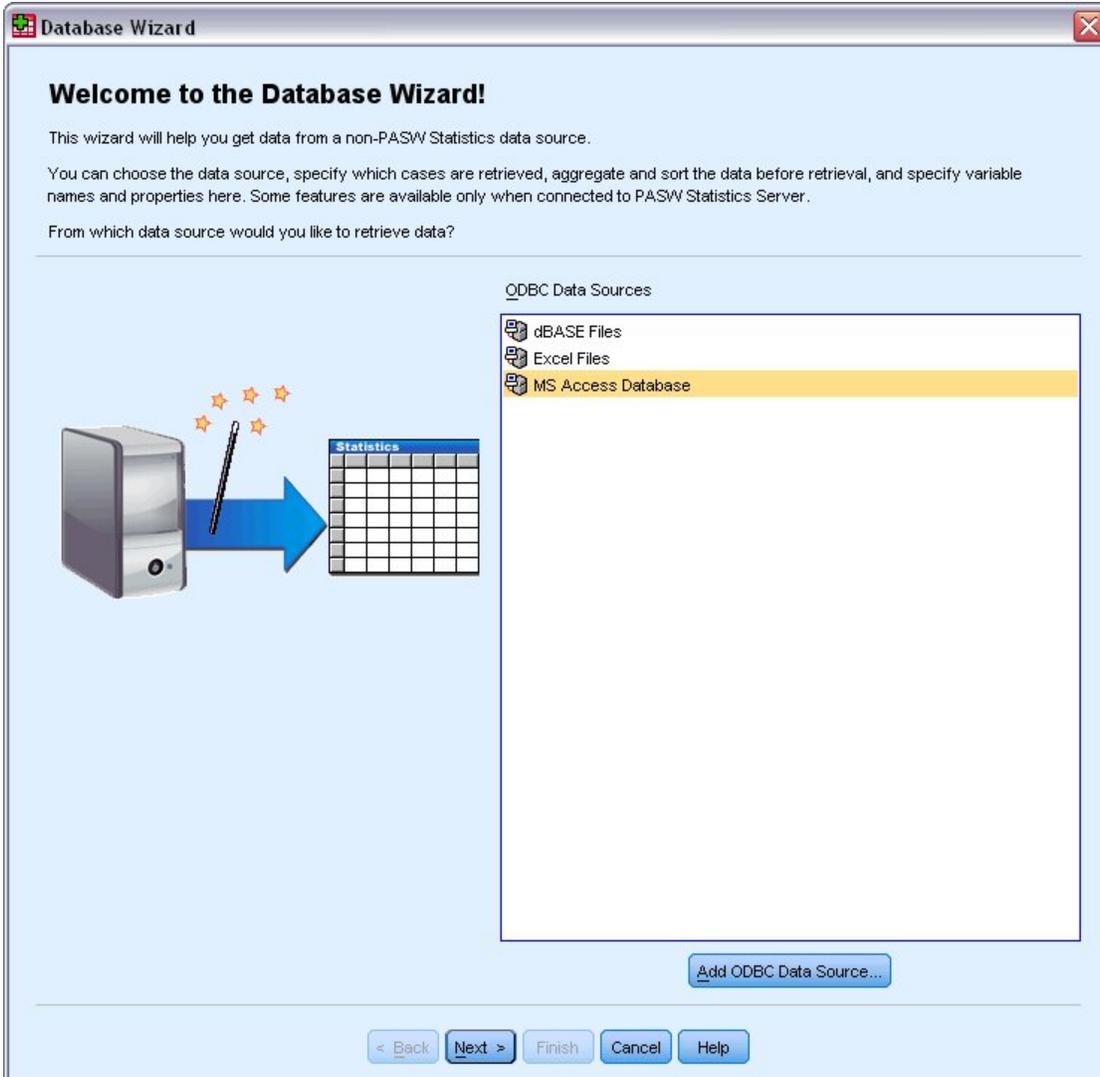


그림 13. 데이터베이스 마법사 시작 대화 상자

2. 데이터 소스 목록에서 **MS Access Database**를 선택하고 다음을 클릭합니다.

참고: 설치에 따라 마법사의 왼쪽에 OLE DB 데이터 소스 목록이 표시될 수도 있지만(Windows 운영체제에만 해당) 이 예제에서는 오른쪽에 표시된 ODBC 데이터 소스 목록을 사용합니다.

3. **찾아보기**를 클릭하여 열려는 Access 데이터베이스 파일을 탐색합니다.
4. *demo.mdb*를 엽니다. 자세한 정보는 91 페이지의 제 10 장 『표본 파일』 주제를 참조하십시오.
5. 로그인 대화 상자에서 **확인**을 클릭합니다.

다음 단계에서는 가져올 표와 변수를 지정할 수 있습니다.

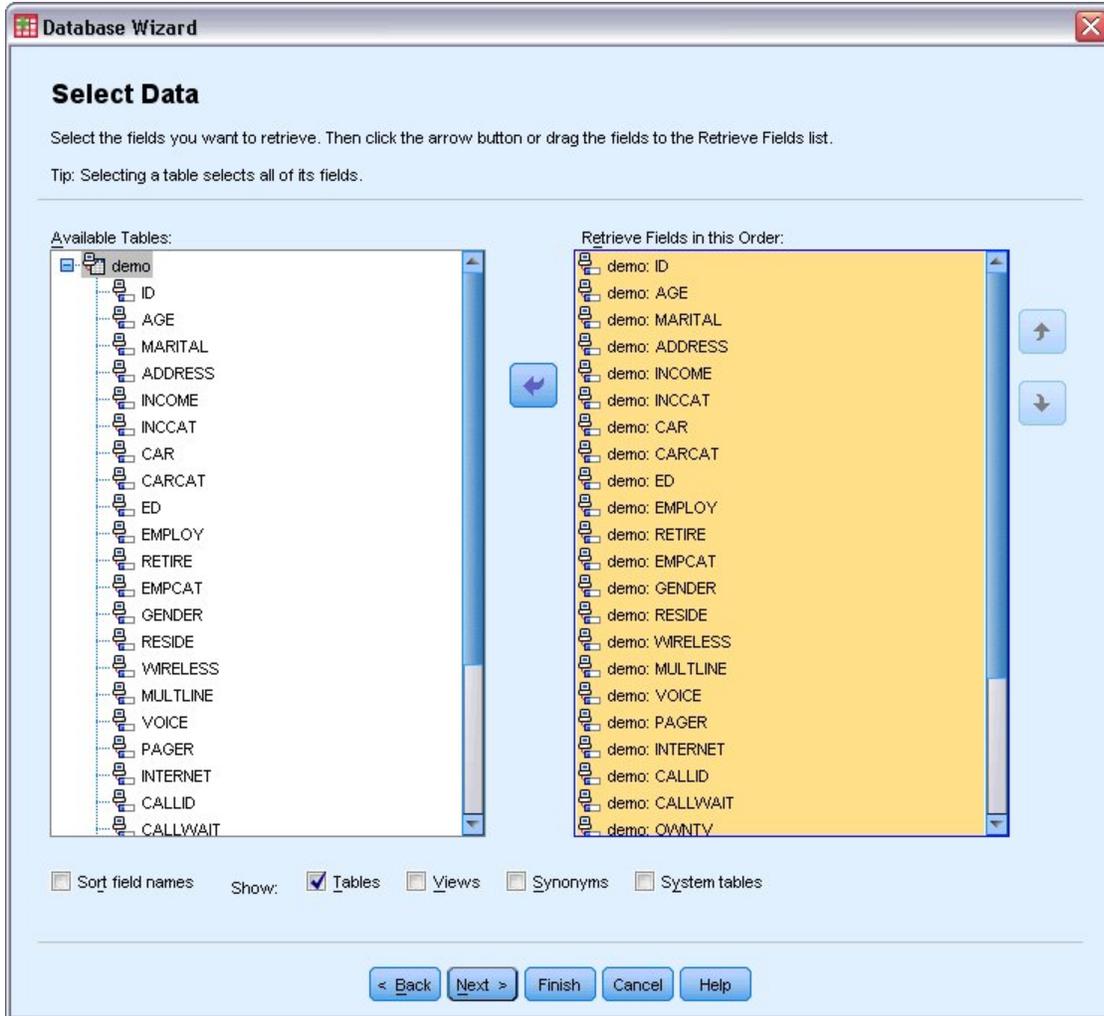


그림 14. 데이터 선택 단계

6. 전체 **demo** 표를 이 순서로 필드 검색 목록으로 끕니다.
7. 다음을 클릭합니다.

다음 단계에서는 가져올 레코드(케이스)를 선택할 수 있습니다.

일부 케이스만 가져오려면 케이스서브세트(예: 30세 이상의 남자)를 가져오거나 데이터 소스에서 무작위로 케이스 표본을 가져올 수 있습니다. 큰 데이터 소스의 경우 케이스 수를 작은 대표 표본으로 제한하여 처리 시간을 줄일 수 있습니다.

8. 계속하려면 다음을 클릭합니다.

필드 이름은 변수 이름을 작성하는 데 사용됩니다. 필요한 경우 이름은 유효한 변수 이름으로 변환됩니다. 원래 필드 이름은 변수 레이블로 유지됩니다. 데이터베이스를 가져오기 전에 변수 이름을 변경할 수도 있습니다.

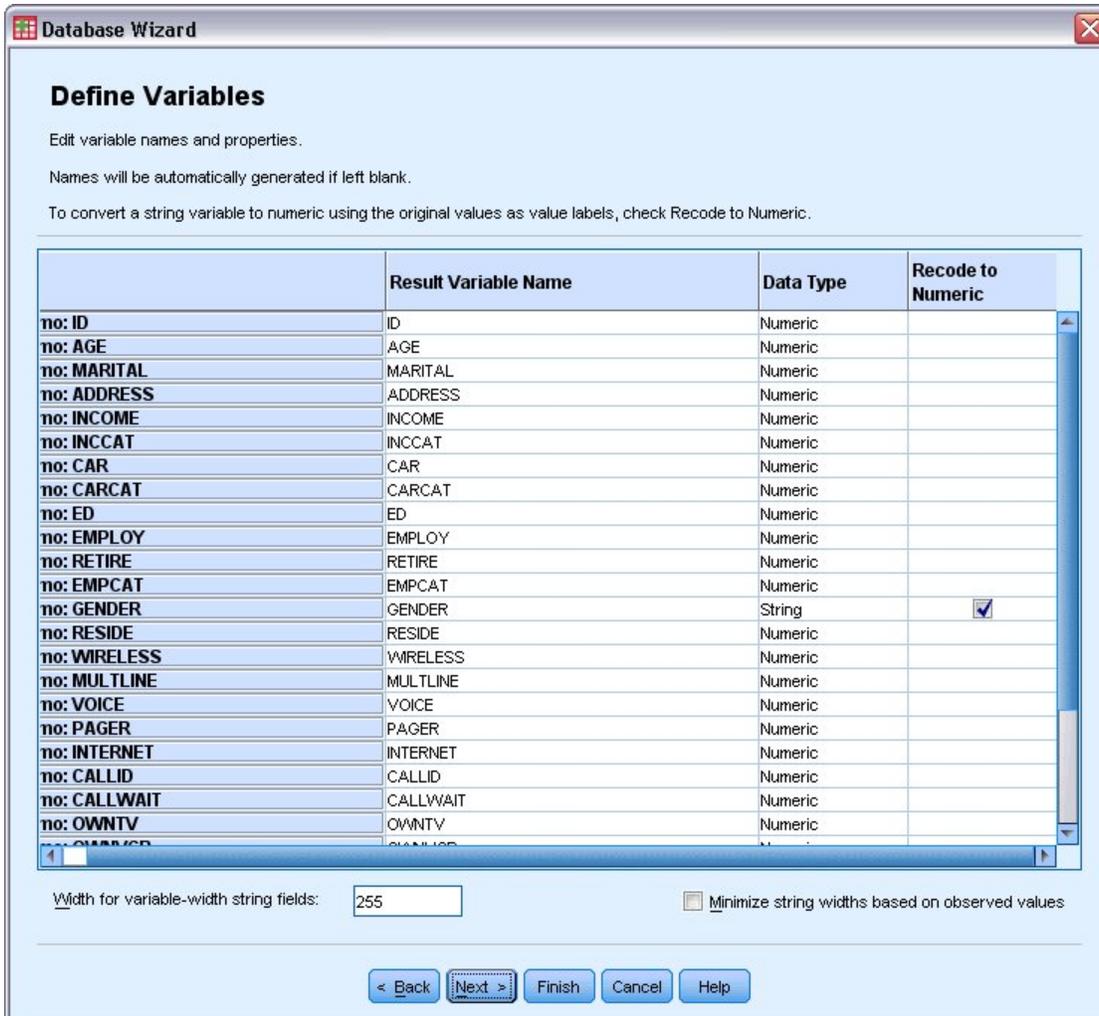


그림 15. 변수 정의 단계

9. 성별 필드의 숫자로 코딩변경 셀을 선택합니다. 이 옵션은 문자열 변수를 정수변수로 변환하고 원래 값을 새 변수의 값 레이블로 유지합니다.
10. 계속하려면 다음을 클릭합니다.  
 데이터베이스 마법사에서 선택한 항목에 따라 작성된 SQL 문이 결과 단계에 나타납니다. 이 SQL 문을 지금 실행하거나 나중에 사용할 수 있도록 저장할 수 있습니다.
11. 데이터를 가져오려면 마침을 클릭합니다.

가져오도록 선택한 Access 데이터베이스의 모든 데이터를 이제 데이터 편집기에서 사용할 수 있습니다.

## 텍스트 파일에서 데이터 읽기

텍스트 파일은 다른 하나의 공통 데이터 소스를 나타냅니다. 많은 스프레드시트 프로그램과 데이터베이스에서 해당 내용을 여러 텍스트 파일 형식 중 하나로 저장할 수 있습니다. 콤마나 탭으로 구분된 파일은 콤마나 탭을 사용하여 각 변수를 나타내는 데이터의 행을 참조합니다. 이 예에서 데이터는 탭으로 구분됩니다.

1. 메뉴에서 다음을 선택합니다.

파일 > 데이터 가져오기 > 텍스트 데이터

2. Samples\English 폴더로 이동하여 demo.txt를 선택합니다.

텍스트 가져오기 마법사는 지정한 텍스트 파일에 대한 해석 방법을 정의하는 프로세스를 안내합니다.

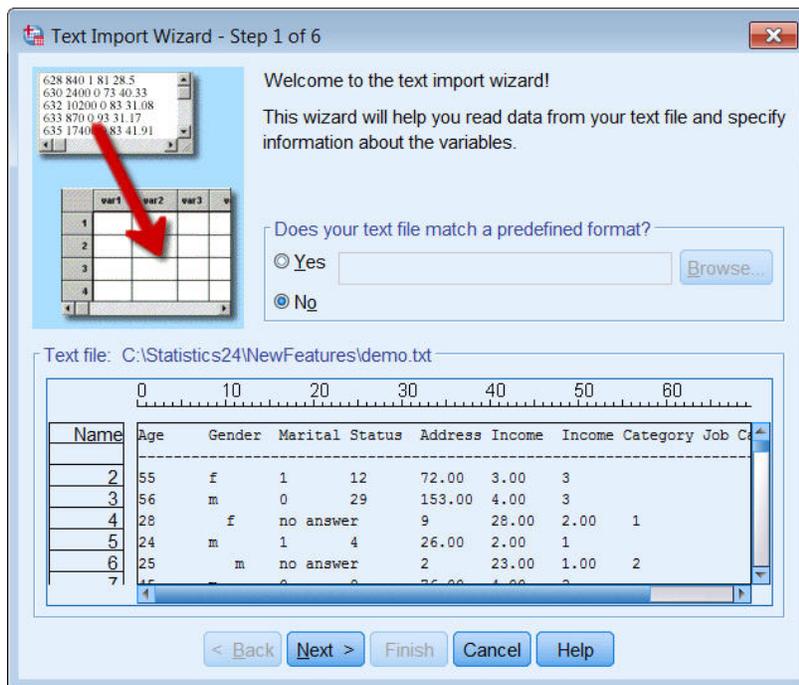


그림 16. 텍스트 가져오기 마법사: 6단계 중 1단계

3. 1단계에서는 사전 정의한 형식을 선택하거나 마법사로 새 형식을 작성할 수 있습니다. 아니오를 선택합니다.
4. 계속하려면 다음을 클릭합니다.

앞에서 언급한 대로 이 파일에서는 탭으로 구분된 형식을 사용합니다. 또한, 변수 이름은 이 파일의 맨 위 행에 정의됩니다.

5. 마법사의 2단계에서 구분자에 의한 배열을 선택하여 파일의 구조가 구분 형식 구조임을 표시합니다.
6. 예를 선택하여 파일 맨 위에 변수 이름이 포함됨을 표시합니다.

7. 계속하려면 다음을 클릭합니다.
8. 변수 이름이 첫 번째 행에 있으므로 3단계에서 데이터의 첫 번째 케이스가 시작하는 행 번호에 대해 2를 입력하십시오.
9. 이 단계의 나머지 옵션은 기본값을 유지하고 계속하려면 다음을 클릭합니다.

4단계의 데이터 미리보기를 통해 파일을 제대로 읽고 있는지 신속하게 확인할 수 있습니다.

10. 탭을 선택하고 구분자의 다른 옵션을 선택 해제합니다. 파일에 공백이 있으므로 기본적으로 공백이 선택됩니다. 이 파일의 경우, 공백은 데이터 값의 일부이고 구분자가 아닙니다. 파일을 올바르게 읽으려면 공백을 선택 취소해야 합니다.
11. 문자열 값의 선행 공백 제거를 선택합니다. 문자열 값의 맨 앞에 있는 공백은 표현식에서 문자열 값이 평가되는 방법에 영향을 미칩니다. 이 파일에서 *Gender*의 일부 값은 값의 일부가 아닌 선행 공백을 포함합니다. 그러한 공백을 제거하지 않는 경우 " f" 값은 "f"와 다른 값으로 처리됩니다.

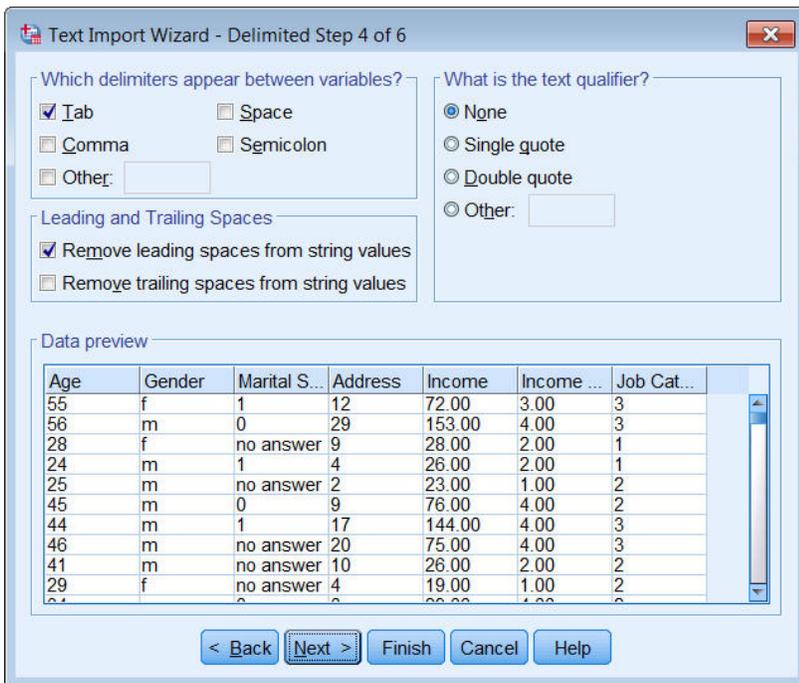


그림 17. 텍스트 가져오기 마법사: 6단계 중 4단계

12. 계속하려면 다음을 클릭합니다.

변수 이름이 이름 규칙을 따르도록 수정되므로 5단계에서는 적합하지 않은 이름을 편집할 수 있는 기회를 제공합니다.

여기서 데이터 유형을 정의할 수도 있습니다. 예를 들어, *Income*을 달러 통화 형식으로 변경할 수 있습니다.

데이터 유형 변경:

13. 데이터 미리보기에서 *Income*을 선택합니다.

14. 데이터 형식 드롭다운 목록에서 달러를 선택합니다.

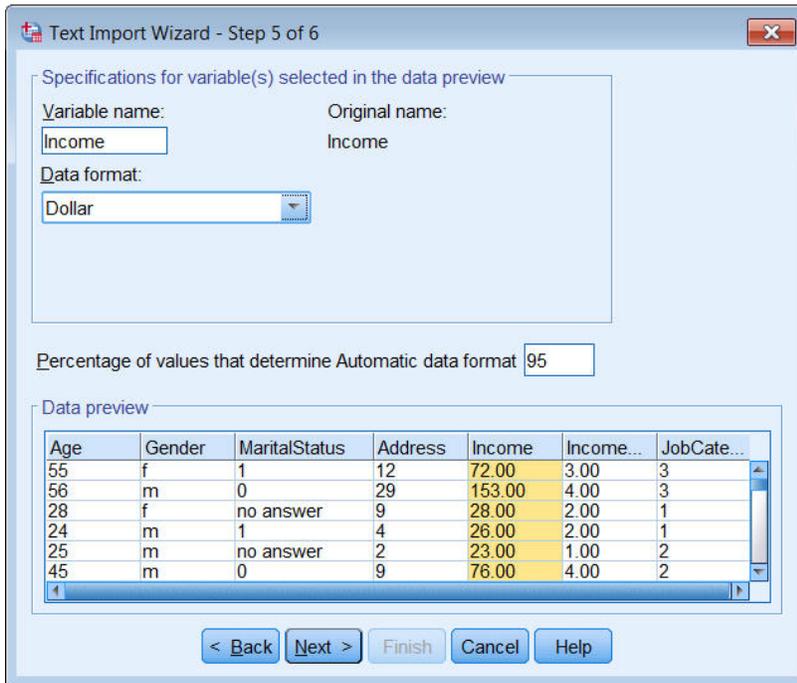


그림 18. 데이터 형식 바꾸기

*MaritalStatus* 변수에는 문자열 값과 숫자값이 모두 포함됩니다. 값 중 5% 미만이 문자열입니다. 자동 데이터 형식을 결정하는 값의 퍼센트에 기본 설정 95%를 사용하는 경우 변수는 숫자로 처리되고 문자열 값은 시스템 결측값으로 설정됩니다. 퍼센트 값을 충족하는 데이터 형식이 없으면 변수는 문자열 변수로 처리됩니다. 설정을 100으로 변경하면 모든 값이 보존되지만 모든 숫자값이 문자열로 처리됩니다.

15. 계속하려면 다음을 클릭합니다.

16. 데이터를 가져오려면 마지막 단계의 선택을 기본값으로 유지하고 마침을 클릭합니다.



---

## 제 3 장 데이터 편집기 사용

데이터 편집기는 활성 데이터 파일의 내용을 표시합니다. 데이터 편집기의 정보는 변수와 케이스로 구성됩니다.

- 데이터 보기에서 열은 변수를 나타내고 행은 케이스(관측값)를 나타냅니다.
- 변수 보기에서 각 행은 변수이고 각 열은 해당 변수와 관련된 속성입니다.

변수는 컴파일한 다른 유형의 데이터를 나타내는 데 사용되며 일반적으로 설문조사에 의한 유추법이 적용됩니다. 설문조사에서 각 질문에 대한 반응이 변수에 해당합니다. 변수는 숫자, 문자열, 통화 및 날짜를 비롯하여 다양한 유형으로 표시됩니다.

---

### 숫자 데이터 입력

데이터 편집기에 데이터를 입력할 수 있으며 작은 데이터 파일을 사용하거나 큰 데이터 파일을 약간씩 편집하는 경우에 유용합니다.

1. 데이터 편집기 창의 아래에서 **변수 보기** 탭을 클릭합니다.

사용할 변수를 정의해야 합니다. 이 경우에는 세 가지 변수인 연령, 결혼 상태 및 가구소득만 필요합니다.

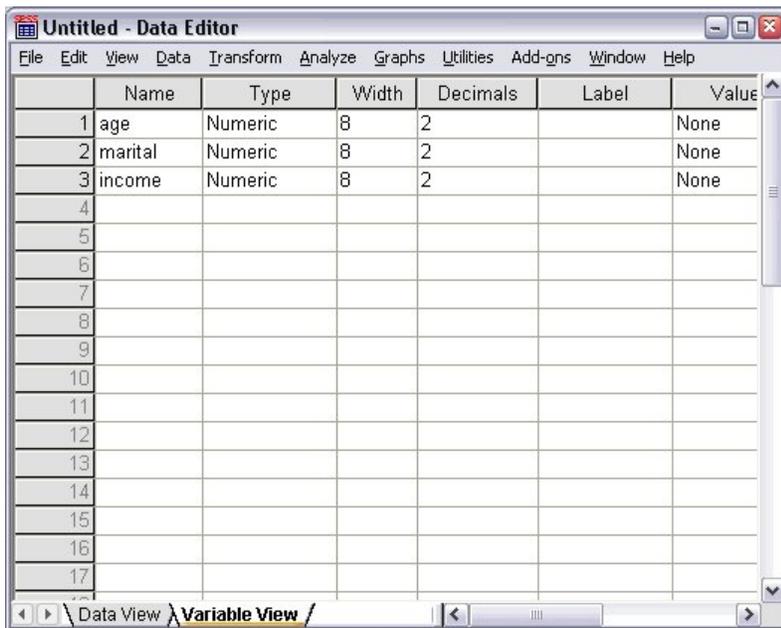


그림 19. 변수 보기의 변수 이름

2. 첫 번째 열의 첫 번째 행에 연령을 입력합니다.

3. 두 번째 행에 결혼 상태를 입력합니다.
4. 세 번째 행에 가구소득을 입력합니다.

새 변수에는 자동으로 숫자 데이터 유형이 제공됩니다.

변수 이름을 입력하지 않을 경우 고유한 이름이 자동으로 작성됩니다. 그러나 자동으로 작성되는 이름은 해당 변수를 설명하지 않으므로 큰 데이터 파일의 경우 직접 입력하는 것이 좋습니다.

5. 데이터 보기 탭을 클릭하여 데이터 입력을 계속합니다.

변수 보기에 입력한 이름이 이제 데이터 보기의 처음 세 개 열에서 머리말이 됩니다.

첫 번째 열부터 시작하여 첫 번째 행에 데이터 입력을 시작합니다.

The screenshot shows the 'Untitled - Data Editor' window in SPSS. The menu bar includes File, Edit, View, Data, Transform, Analyze, Graphs, Utilities, Add-ons, Window, and Help. The active tab is 'Data View'. The data table has the following structure:

	age	marital	income	var	var	var
1	55.00	1.00	72000.00			
2	53.00	.00	153000.0			
3						
4						
5						
6						
7						
8						
9						
10						
11						
12						
13						
14						
15						
16						

그림 20. 데이터 보기에 입력한 값

6. 연령 열에 55를 입력합니다.
7. 결혼 상태 열에 1을 입력합니다.
8. 가구소득 열에 72000을 입력합니다.
9. 커서를 첫 번째 열의 두 번째 행으로 이동하여 다음 개체의 데이터를 추가합니다.
10. 연령 열에 53을 입력합니다.
11. 결혼 상태 열에 0을 입력합니다.
12. 가구소득 열에 153000을 입력합니다.

연령 및 결혼 상태 열은 정수 값을 가져야 하지만 기본적으로 소수점 형식으로 표시됩니다. 이러한 변수에서 소수점을 감추려면 다음과 같이 하십시오.

13. 데이터 편집기 창의 아래에서 변수 보기 탭을 클릭합니다.

14. 연령 행의 소수점이하자리 열에 0을 입력하여 소수점이하자리를 감춥니다.
15. 결혼 상태 행의 소수점이하자리 열에 0을 입력하여 소수점이하자리를 감춥니다.

## 문자열 데이터 입력

텍스트의 문자와 같은 비슷자 데이터를 데이터 편집기에 입력할 수도 있습니다.

1. 데이터 편집기 창의 아래에서 **변수 보기** 탭을 클릭합니다.
2. 비어 있는 첫 번째 행의 첫 번째 셀에 변수 이름으로 성별을 입력합니다.
3. 항목 옆의 유형 셀을 클릭합니다.
4. 유형 셀 오른쪽의 단추를 클릭하여 변수 유형 대화 상자를 엽니다.
5. **문자열**을 선택하여 변수 유형을 지정합니다.
6. **확인**을 클릭하여 선택을 저장하고 데이터 편집기로 돌아갑니다.

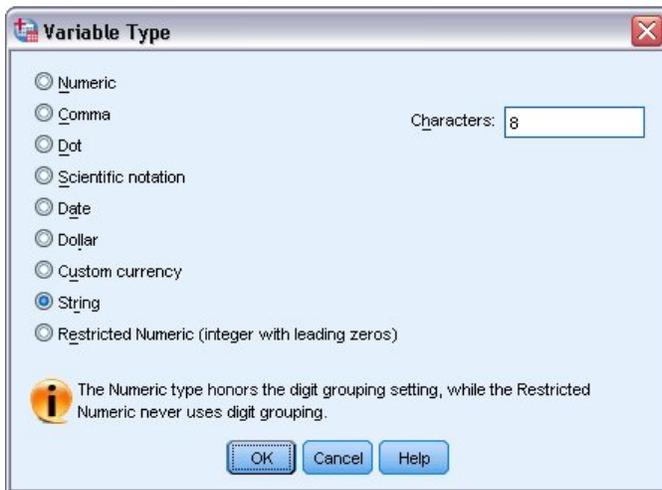


그림 21. 변수 유형 대화 상자

## 데이터 정의

데이터 유형 정의와 함께 변수 이름과 데이터 값에 대해 변수 설명 레이블 및 값 레이블도 정의할 수 있습니다. 이러한 설명 레이블은 통계 보고서와 도표에 사용됩니다.

### 변수 레이블 추가

레이블에서는 변수 이름을 서술합니다. 대부분 변수 이름을 길게 풀어 쓴 것입니다. 레이블은 255바이트를 넘지 않아야 합니다. 이러한 레이블은 결과에서 다른 변수를 식별하는 데 사용됩니다.

1. 데이터 편집기 창의 아래에서 **변수 보기** 탭을 클릭합니다.
2. 연령 행의 레이블 열에 반응자 연령을 입력합니다.

3. 결혼 상태 행의 레이블 열에 결혼 상태를 입력합니다.
4. 가구소득 행의 레이블 열에 가구소득을 입력합니다.
5. 성별 행의 레이블 열에 성별을 입력합니다.

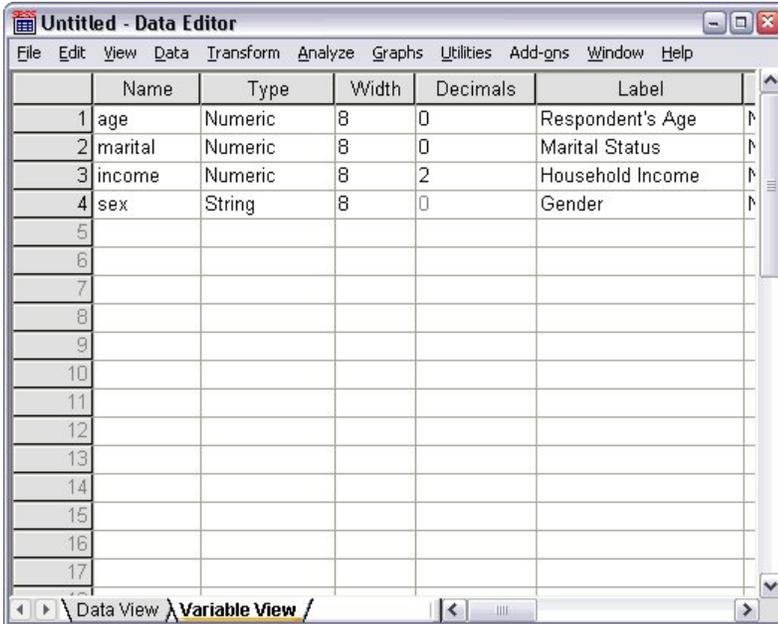


그림 22. 변수 보기에 입력된 값 레이블

## 변수 유형 및 형식 바꾸기

유형 열에는 각 변수의 현재 데이터 유형이 표시됩니다. 가장 일반적인 데이터 유형은 숫자와 문자열이지만 기타 여러 가지 형식이 지원됩니다. 현재 데이터 파일에서 가구소득 변수는 숫자 유형으로 정의됩니다.

1. 가구소득 행에서 유형 셀을 클릭한 다음 셀 오른쪽의 단추를 클릭하여 변수 유형 대화 상자를 엽니다.
2. 달러를 선택합니다.

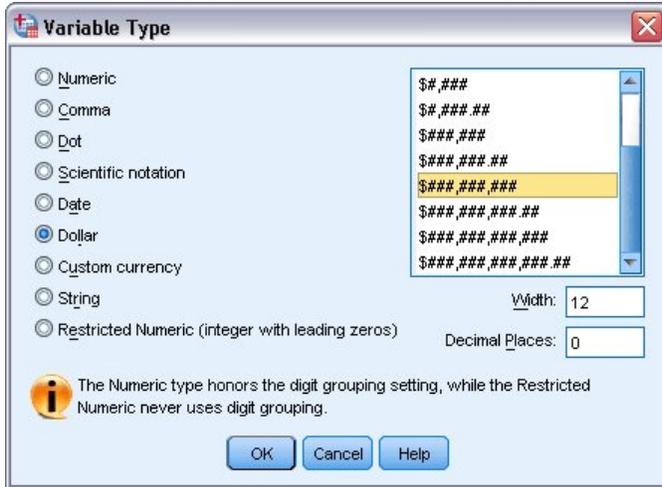


그림 23. 변수 유형 대화 상자

현재 선택한 데이터 유형에 대한 형식 지정 옵션이 표시됩니다.

3. 이 예제에서는 통화 형식으로 `$###,###,###`을 선택합니다.
4. 확인을 클릭하여 변경 내용을 저장합니다.

## 값 레이블 추가

값 레이블의 변수값을 문자열 레이블에 매핑할 수 있습니다. 이 예제에서는 결혼 상태 변수에 두 가지 값이 허용됩니다. 값이 0이면 개체가 미혼이며 값이 1이면 기혼임을 의미합니다.

1. 결혼 상태 행에서 값 셀을 클릭한 다음 셀 오른쪽의 단추를 클릭하여 값 레이블 대화 상자를 엽니다.

값은 실제 숫자값입니다.

값 레이블은 지정된 숫자값에 적용되는 문자열 레이블입니다.

2. 변수값 필드에 0을 입력합니다.
3. 레이블 필드에 미혼을 입력합니다.
4. 추가를 클릭하여 이 레이블을 목록에 추가합니다.

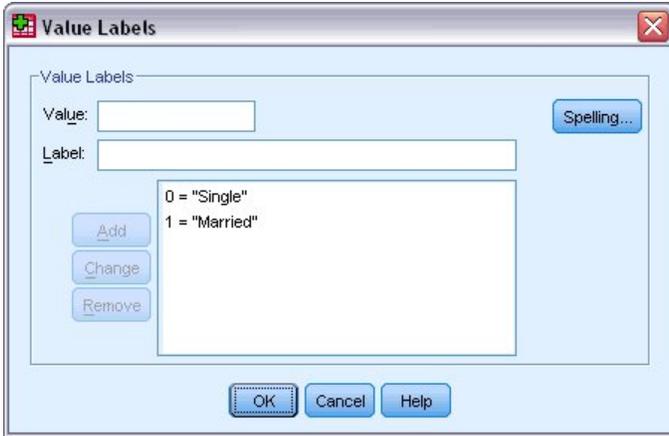


그림 24. 값 레이블 대화 상자

5. 변수값 필드에 1을 입력한 다음 레이블 필드에 기혼을 입력합니다.
6. 추가를 클릭한 다음 확인을 클릭하여 변경 내용을 저장하고 데이터 편집기로 돌아갑니다.

이러한 레이블은 또한 데이터 보기에도 표시될 수 있으며 이 경우 데이터를 더 읽기 쉽게 만드는 데 도움이 됩니다.

7. 데이터 편집기 창의 아래에서 데이터 보기 탭을 클릭합니다.
8. 메뉴에서 다음을 선택합니다.

#### 보기 > 값 레이블

이제 데이터 편집기에 값을 입력할 때 목록에 레이블이 표시됩니다. 이 설정은 유효한 반응을 제안하고 반응을 구체적으로 서술하는 데 유용합니다.

값 레이블 메뉴 항목이 이미 활성화되어 있는 경우(항목 옆에 체크 표시가 있음) 값 레이블을 다시 선택하면 값 레이블 표시가 해제됩니다.

### 결측 데이터 처리

일반적으로 결측 데이터나 유효하지 않은 데이터가 발생하는 이유는 대개 비슷하기 때문에 분석 시 무시할 수 없습니다. 설문조사 반응자가 특정 질문에 대답하기를 거부하거나, 답을 알지 못하거나, 예상치 않은 형식으로 대답하기 때문입니다. 이러한 데이터를 필터링 또는 식별하지 않으면 분석 결과가 정확하지 않을 수 있습니다.

숫자 데이터의 경우 빈 데이터 필드나 잘못된 항목을 포함하는 필드는 하나의 마침표로 식별 가능한 시스템-결측값으로 변환됩니다.

결측이 발생하는 이유는 분석에 중요할 수 있습니다. 예를 들어, 질문에 대답하는 것을 거부한 반응자와 해당 사항이 없기 때문에 질문에 대답하지 않은 반응자를 서로 구별하는 것이 유용할 수 있습니다.

## 숫자변수에 대한 결측값

1. 데이터 편집기 창의 아래에서 **변수 보기** 탭을 클릭합니다.
2. 연령 행에서 결측 셀을 클릭한 다음 셀 오른쪽의 단추를 클릭하여 결측값 대화 상자를 엽니다.

이 대화 상자에서 최대 세 개의 이산형 결측값을 지정하거나 값 범위와 하나의 추가 이산값을 지정할 수 있습니다.



그림 25. 결측값 대화 상자

3. **이산형 결측값**을 선택합니다.
4. 첫 번째 입력란에 999를 입력하고 다른 두 개의 입력란은 비워 둡니다.
5. **확인**을 클릭하여 변경 내용을 저장하고 데이터 편집기로 돌아갑니다.

이제 결측 데이터 값이 추가되었으며 레이블을 해당 값에 적용할 수 있습니다.

6. 연령 행에서 값 셀을 클릭한 다음 셀 오른쪽의 단추를 클릭하여 값 레이블 대화 상자를 엽니다.
7. 변수값 필드에 999를 입력합니다.
8. 레이블 필드에 무반응을 입력합니다.
9. **추가**를 클릭하여 이 레이블을 데이터 파일에 추가합니다.
10. **확인**을 클릭하여 변경 내용을 저장하고 데이터 편집기로 돌아갑니다.

## 문자열 변수에 대한 결측값

문자열 변수에 대한 결측값은 숫자변수에 대한 결측값과 비슷하게 처리됩니다. 그러나 숫자변수와 달리 문자열 변수의 빈 필드는 시스템-결측값으로 지정되지 않습니다. 대신, 이러한 필드는 빈 문자열로 해석됩니다.

1. 데이터 편집기 창의 아래에서 **변수 보기** 탭을 클릭합니다.
2. 성별 행에서 결측 셀을 클릭한 다음 셀 오른쪽의 단추를 클릭하여 결측값 대화 상자를 엽니다.
3. **이산형 결측값**을 선택합니다.
4. 첫 번째 입력란에 NR을 입력합니다.

문자열 변수에 대한 결측값은 대소문자를 구분합니다. 따라서 *nr* 값은 결측값으로 간주되지 않습니다.

5. **확인**을 클릭하여 변경 내용을 저장하고 데이터 편집기로 돌아갑니다.

이제 결측값에 대한 레이블을 추가할 수 있습니다.

6. 성별 행에서 값 셀을 클릭한 다음 단추를 클릭하여 값 레이블 대화 상자를 엽니다.
7. 변수값 필드에 NR을 입력합니다.
8. 레이블 필드에 무반응을 입력합니다.
9. **추가**를 클릭하여 이 레이블을 프로젝트에 추가합니다.
10. **확인**을 클릭하여 변경 내용을 저장하고 데이터 편집기로 돌아갑니다.

---

## 제 4 장 개별 변수의 요약 통계 탐색

이 절에서는 간단한 요약 측도 및 사용해야 하는 통계 유형에 변수 측정 수준이 주는 영향에 대해 설명합니다. 데이터 파일 *demo.sav*를 사용할 것입니다. 자세한 정보는 91 페이지의 제 10 장 『표본 파일』 주제를 참조하십시오.

---

### 측정 수준

데이터 유형별로 측정 수준에 맞는 요약 측정이 필요합니다.

**범주형.** 고유한 값/범주 수가 제한된 데이터(예: 성별 또는 결혼 상태)입니다. **질적 데이터**라고도 합니다. 문자열(문숫자) 데이터 변수 또는 숫자 코드를 사용하여 범주를 나타내는 숫자변수(예: 0 = 미혼, 1 = 기혼)가 범주형 변수에 해당될 수 있습니다. 두 가지 기본 범주형 데이터 유형이 있습니다.

- **명목.** 범주에 순서가 없는 범주형 데이터입니다. 예를 들어, 판매 직업 범주는 마케팅 또는 조사 직업 범주와 비교하여 높고 낮음의 순서를 구분할 수 없습니다.
- **순서.** 범주 간에 의미상의 순서는 있지만 이러한 범주 간의 차이는 측정할 수 없는 범주형 데이터입니다. 예를 들어, 높음, 중간, 낮음과 같이 값의 순서가 있어도 이러한 값들 간의 "차이"는 계산할 수 없습니다.

**척도.** 구간 또는 비율 척도로 측정된 데이터에서 데이터 값은 값의 순서와 차이를 모두 나타냅니다. 예를 들어, \$72,195의 급여는 \$52,398보다 높은 순위이고 두 값의 차이는 \$19,797입니다. 이러한 데이터는 **양적 데이터** 또는 **연속형 데이터**라고도 합니다.

---

### 범주형 데이터에 대한 요약 측도

범주형 데이터의 경우 가장 일반적인 요약 측도는 각 범주의 케이스 수 또는 퍼센트입니다. **최빈값**은 케이스가 가장 많은 범주입니다. 순서 데이터의 경우 **중앙값**(전체 케이스의 절반이 위아래에 해당되는 값)은 범주가 많을 경우 유용한 요약 측도가 될 수도 있습니다.

빈도분석 프로시저는 변수의 각 관측값에 대한 케이스 수와 퍼센트를 모두 표시하는 빈도표를 생성합니다.

1. 메뉴에서 다음을 선택합니다.

분석 > 기술통계량 > 빈도...

**참고:** 이 기능을 사용하려면 Statistics Base Edition이 필요합니다.

2. *PDA* 소유 [*PDA*보유]와 *TV* 소유 [*TV*보유]를 선택하여 변수 목록으로 이동합니다.

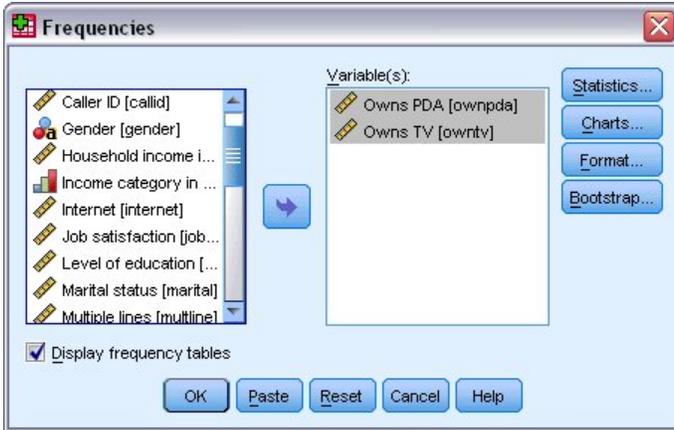


그림 26. 분석을 위해 선택한 범주형 변수

3. **확인**을 클릭하여 프로시저를 실행합니다.

**Frequency Table**

**Owns PDA**

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	No	5093	79.6	79.6	79.6
	Yes	1307	20.4	20.4	100.0
Total		6400	100.0	100.0	

**Owns TV**

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	No	63	1.0	1.0	1.0
	Yes	6337	99.0	99.0	100.0
Total		6400	100.0	100.0	

그림 27. 빈도표

빈도표가 뷰어 창에 표시됩니다. 이 빈도표를 보면 약 20.4%만이 PDA를 소유하고 있지만 거의 모든 사람이 TV(99.0%)를 소유하고 있음을 알 수 있습니다. 이는 TV를 소유하지 않은 소규모 그룹에 대한 자세한 정보를 찾을 때는 유용하지만 그리 적합한 표 형식은 아닙니다.

## 범주형 데이터에 대한 도표

빈도표에 막대도표나 원형 도표로 정보를 그림으로 표시할 수 있습니다.

1. 빈도분석 대화 상자를 다시 엽니다. 두 개의 변수가 선택되어 있어야 합니다.

도구 모음에서 대화 상자 다시 불러오기 단추를 사용하여 최근에 사용한 프로시저로 바로 돌아갈 수 있습니다.



그림 28. 대화 상자 다시 불러오기 단추

2. 도표를 클릭합니다.
3. 막대도표를 선택한 다음 계속을 클릭합니다.
4. 기본 대화 상자에서 확인을 클릭하여 프로시저를 실행합니다.

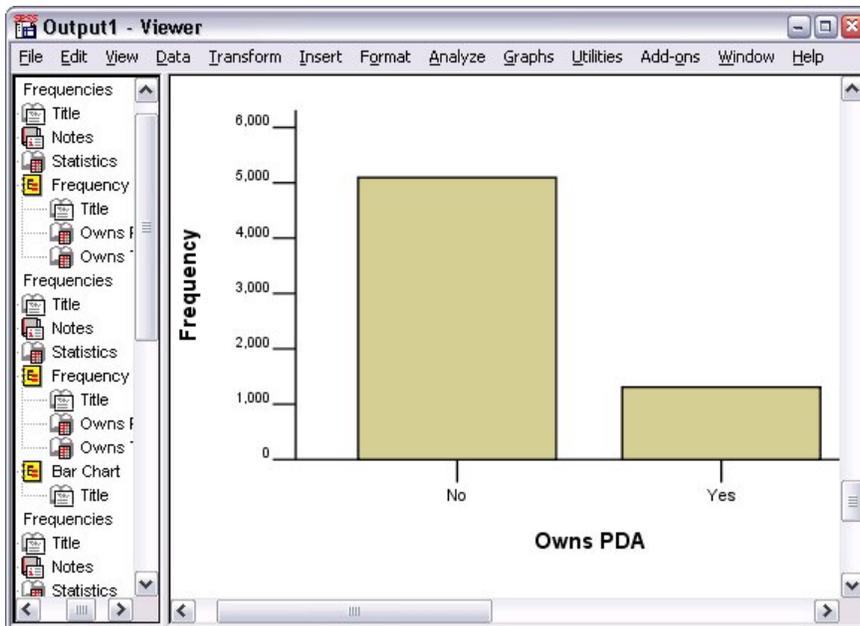


그림 29. 막대도표

빈도표에 나와 있는 정보가 그대로 이 막대도표 형식으로 표시되므로 PDA를 소유하고 있는 사람은 매우 적지만 거의 모든 사람이 TV를 소유하고 있음을 쉽게 알 수 있습니다.

## 척도변수에 대한 요약 측도

다음은 비롯하여 척도변수에 사용할 수 있는 여러 가지 요약 측도가 있습니다.

- **중심 경향 측도.** 가장 일반적인 중심 경향 측도는 **평균**(산술 평균)과 **중앙값**(전체 케이스의 반은 값의 위에, 나머지 반은 아래에 포함시키는 중간값)입니다.
- **산포 측도.** 데이터에 분포되어 있는 변동의 양을 측도하는 통계에는 표준 편차, 최소값, 최대값이 있습니다.

1. 빈도분석 대화 상자를 다시 엽니다.

2. 이전 설정을 지우려면 **재설정**을 클릭합니다.
3. 가구소득 [단위:1000] [가구소득]을 선택하여 변수 목록으로 이동합니다.
4. **통계**를 클릭합니다.
5. **평균, 중앙값, 표준 편차, 최소값, 최대값**을 선택합니다.
6. **계속**을 클릭합니다.
7. 기본 대화 상자에서 **빈도표 출력**을 선택 해제합니다. 빈도표는 데이터 파일에 있는 케이스 수만큼의 고유한 값을 가지고 있기 때문에 척도변수에는 유용하지 않습니다.
8. **확인**을 클릭하여 프로시저를 실행합니다.

빈도 통계표가 뷰어 창에 표시됩니다.

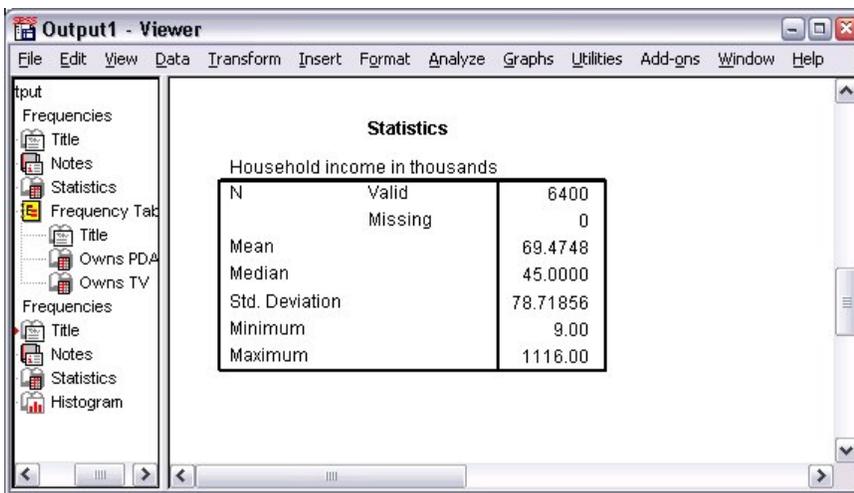


그림 30. 빈도 통계표

이 예제에서는 평균과 중앙값의 차이가 매우 큼니다. 평균이 중앙값보다 거의 25,000 이상 큰데 이는 값이 정상적으로 분포되지 않았음을 나타냅니다. 히스토그램을 사용하여 이러한 분포 상황을 시각적으로 확인할 수 있습니다.

### 척도변수에 대한 히스토그램

1. 빈도분석 대화 상자를 다시 엽니다.
2. **도표**를 클릭합니다.
3. **히스토그램과 정규 곡선 표시**를 선택합니다.
4. **계속**을 클릭한 다음 기본 대화 상자의 **확인**을 클릭하여 프로시저를 실행합니다.

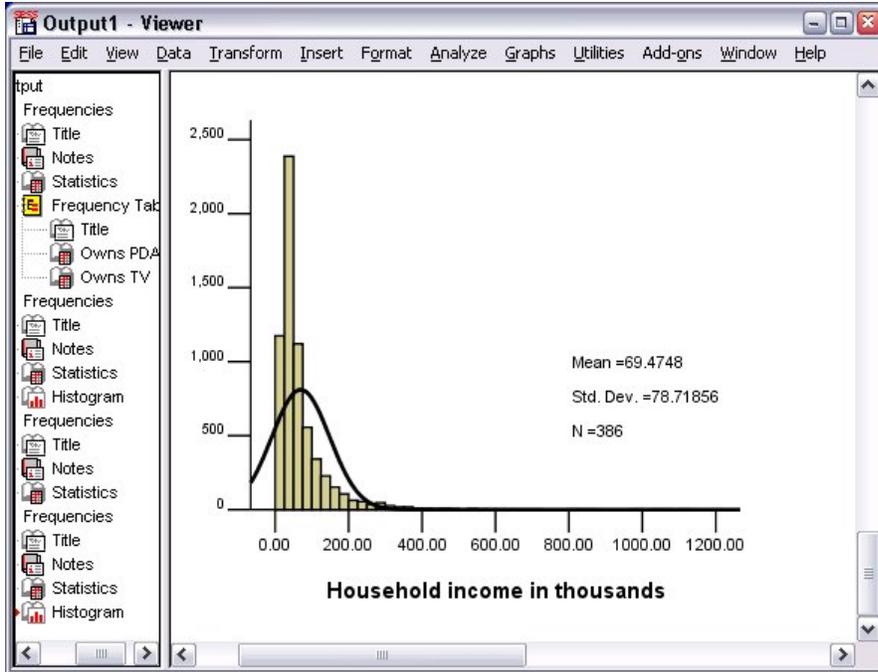


그림 31. 히스토그램

대부분의 케이스가 100,000 아래로 떨어져 척도의 아래쪽 끝에 군집되어 있지만 500,000 범위를 초과하는 케이스가 몇 개 있습니다. 이는 매우 적은 수이기 때문에 히스토그램을 수정해야 볼 수 있습니다. 이러한 소수 케이스의 상한값이 평균에는 중요한 영향을 주지만 중앙값에는 거의 영향을 주지 않으므로 이 예제에서는 중앙값이 중심 경향 지표로 적합합니다.



---

## 제 5 장 도표 작성 및 편집

다양한 도표 유형을 작성하고 편집할 수 있습니다. 이 장에서 막대도표를 작성하고 편집하게 됩니다. 모든 도표 유형에 원리를 적용할 수 있습니다.

---

### 도표 작성 기본 사항

도표 작성 기본 사항을 설명하기 위해 여러 수준의 직업만족도별 평균 소득에 관한 막대도표를 작성합니다. 이 예제에서는 데이터 파일 *demo.sav*를 사용합니다. 자세한 정보는 91 페이지의 제 10 장 『표본 파일』 주제를 참조하십시오.

1. 메뉴에서 다음을 선택합니다.

그래프 > 도표 작성기...

도표 작성기 대화 상자는 도표를 작성하는 동안 도표 모양을 미리 볼 수 있는 대화형 창입니다.

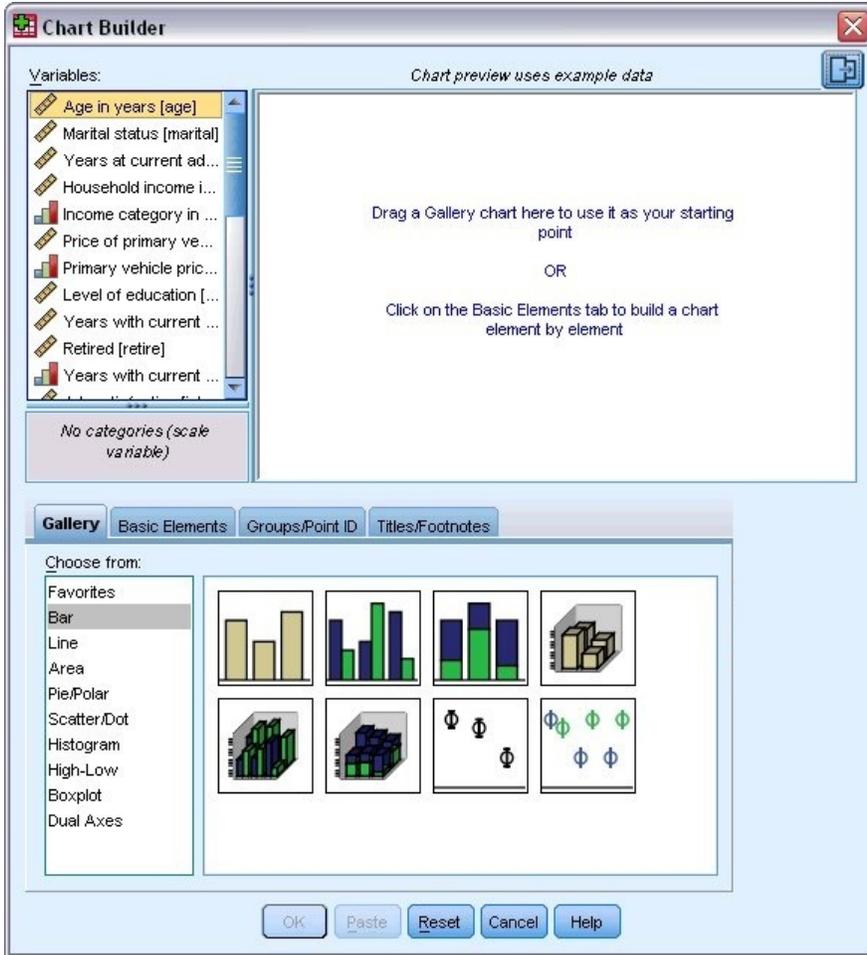


그림 32. 도표 작성기 대화 상자

## 도표 작성기 갤러리 사용

1. 갤러리 탭이 선택되어 있지 않은 경우 갤러리 탭을 클릭합니다.

갤러리에는 여러 가지 사전 정의된 도표가 도표 유형별로 구성되어 있습니다. 기본 요소 탭도 스크래치에서 도표를 작성할 수 있는 기본 요소(예: 축, 그래픽 요소 등)를 제공하지만 갤러리를 사용하는 것이 더 쉽습니다.

2. 막대가 선택되어 있지 않은 경우 막대를 클릭합니다.

갤러리에 사용 가능한 막대도표를 나타내는 아이콘이 대화 상자에 나타납니다. 그림을 보면 특정 도표 유형을 쉽게 확인할 수 있습니다. 추가 정보가 필요한 경우 아이콘 위에 커서를 일시 중지하여 도표의 도구 팁 설명을 표시할 수도 있습니다.

3. 갤러리 위의 넓은 영역인 "캔버스"로 단순 막대도표 아이콘을 끕니다. 캔버스에 도표 작성기의 미리보기가 표시됩니다. 도표를 그리는 데 사용하는 데이터는 실제 데이터가 아니라 예제 데이터입니다.

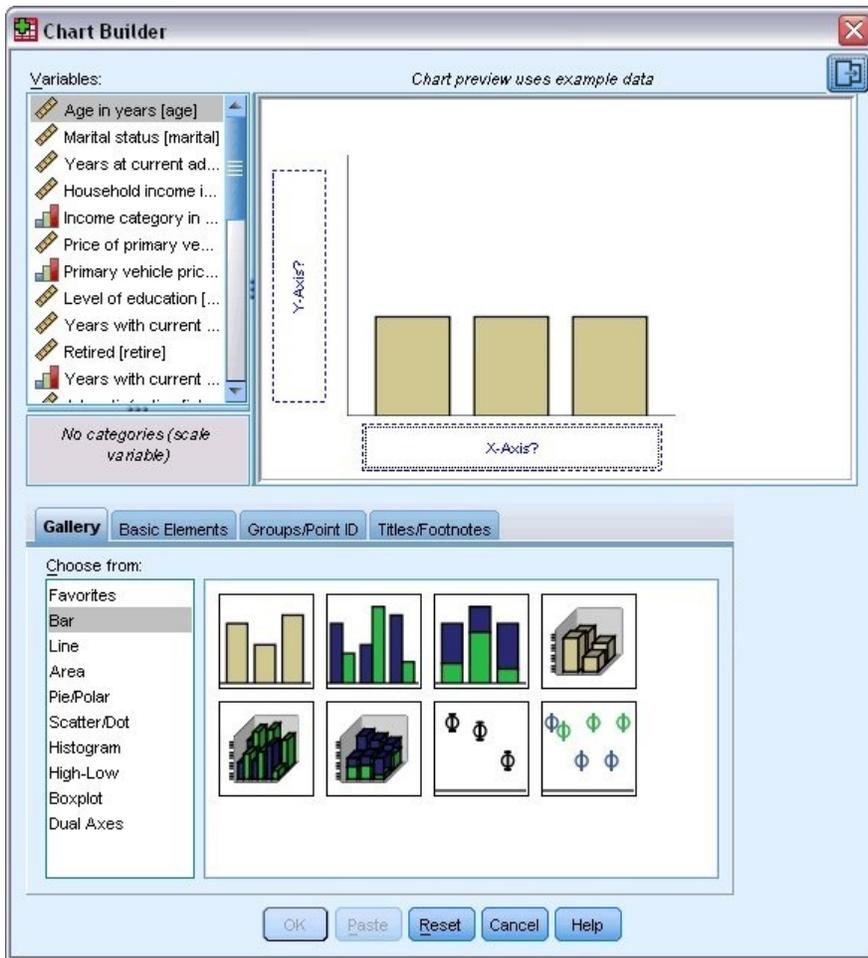


그림 33. 도표 작성기 캔버스의 막대도표

## 변수 및 통계 정의

캔버스에 도표가 있지만 막대 높이를 제어하고 각 막대에 해당하는 변수 범주를 지정할 변수 또는 통계가 없으므로 이 도표는 완전한 것이 아닙니다. 변수 또는 통계가 없는 도표는 작성할 수 없습니다. 캔버스의 왼쪽에 있는 변수 목록에서 변수를 끌어서 추가합니다.

도표 작성기에서는 변수의 측정 수준이 중요합니다.  $x$ 축에는 직업만족도 변수를 사용합니다. 그러나 변수 옆에 있는 눈금자 모양의 아이콘은 해당 측정 수준이 척도로 정의되어 있음을 나타냅니다. 올바른 도표를 작성하려면 범주형 측정 수준을 사용해야 합니다. 이전으로 돌아가 변수 보기에서 측정 수준을 바꾸는 대신 도표 작성기에서 일시적으로 측정 수준을 변경할 수 있습니다.

1. 변수 목록에서 직업만족도를 마우스 오른쪽 단추로 클릭하고 순서를 선택합니다. 직업만족도의 범주는 만족도 수준에 따라 순위를 정할 수 있으므로 순서가 적절한 측정 수준입니다. 측정 수준을 변경하면 아이콘도 따라서 바뀝니다.
2. 이제 직업만족도를 변수 목록에서  $x$ 축 끌어놓기 영역으로 끕니다.

y축 끌어놓기 영역의 기본값은 개수 통계입니다. 그러나 원하는 경우 퍼센트 또는 평균과 같은 다른 통계로 쉽게 바꿀 수 있습니다. 이 예제에서는 이러한 통계를 사용하지 않지만 나중에 통계를 변경해야 하는 경우를 대비하여 해당 프로세스를 검토해 봅니다.

3. 도표 작성기의 세로 막대에서 요소 특성 탭을 클릭합니다. 세로 막대가 표시되지 않으면 도표 작성기의 오른쪽 상단 코너에 있는 단추를 클릭하여 세로 막대를 표시하십시오.



그림 34. 요소 특성

요소 특성에서는 여러 도표 요소의 특성을 변경할 수 있습니다. 이러한 요소에는 막대도표의 막대와 같은 그래픽 요소 및 도표의 축이 포함됩니다. 다음 특성 편집 목록에서 관련된 특성을 변경할 요소 하나를 선택합니다. 또한 목록의 오른쪽에 있는 빨간색 X 표시에 유의합니다. 이 단추는 캔버스에서 그래픽 요소를 삭제합니다. 막대1을 선택했으므로 표시된 특성이 그래픽 요소, 특히 막대 그래픽 요소에 적용됩니다.

통계 드롭다운 목록에서는 사용 가능한 특정 통계를 보여 줍니다. 일반적으로 모든 도표 유형에 동일한 통계를 사용할 수 있습니다. 그러나 일부 통계는 y축 끌어놓기 영역에 변수가 있어야 한다는 점에 유의해야 합니다.

4. 가구소득 [단위:1000]을 변수 목록에서 y축 끌어놓기 영역으로 끕니다. y축 변수는 척도형이고 x축 변수는 범주형(순서는 범주형 측정 수준의 한 유형)이므로 y축 끌어놓기 영역의 기본값은 평균 통계입니다. 원하는 변수 및 통계이므로 요소 특성을 바꿀 필요가 없습니다.

## 텍스트 추가

도표에 제목 및 꼬리말을 추가할 수도 있습니다.

1. 제목/꼬리말 탭을 클릭합니다.
2. 제목 1을 선택합니다.

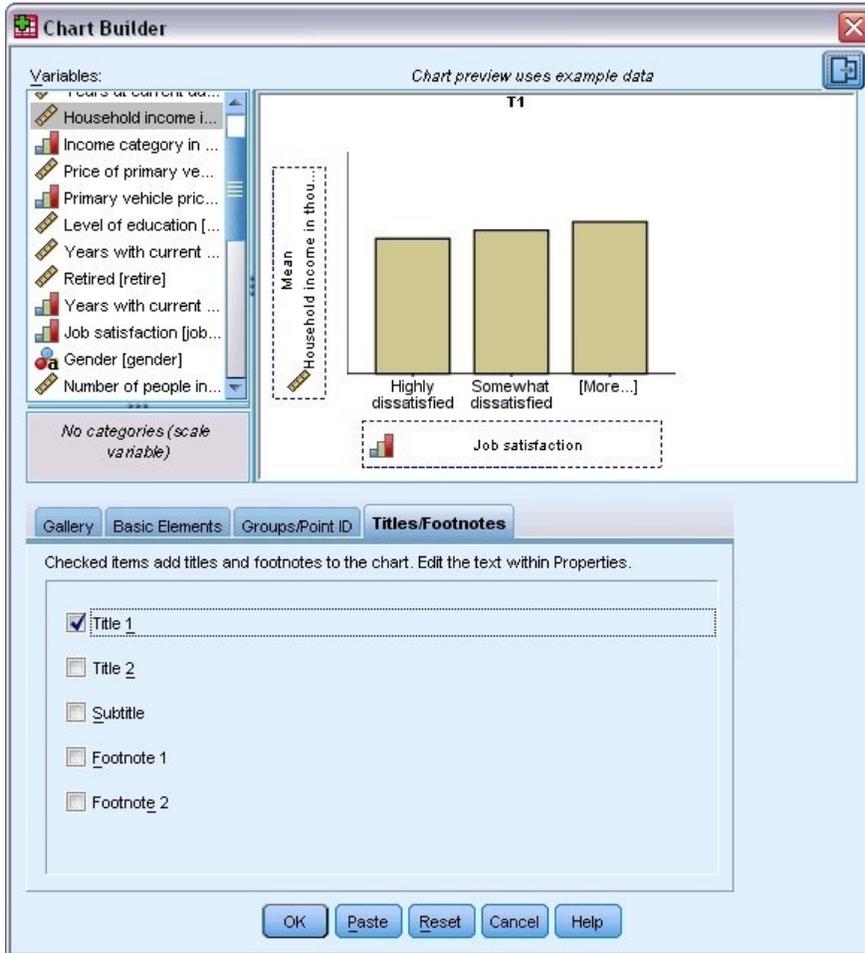


그림 35. 캔버스 분할창에 표시된 제목 1

캔버스에 T1이라는 레이블과 함께 제목이 표시됩니다.

3. 요소 특성 탭의 다음 특성 편집 목록에서 제목 1을 선택합니다.
4. 내용 입력란에 직업만족도별 소득을 입력합니다. 이것은 제목이 표시되는 텍스트입니다.

## 도표 작성

1. 확인을 클릭하여 막대도표를 작성합니다.

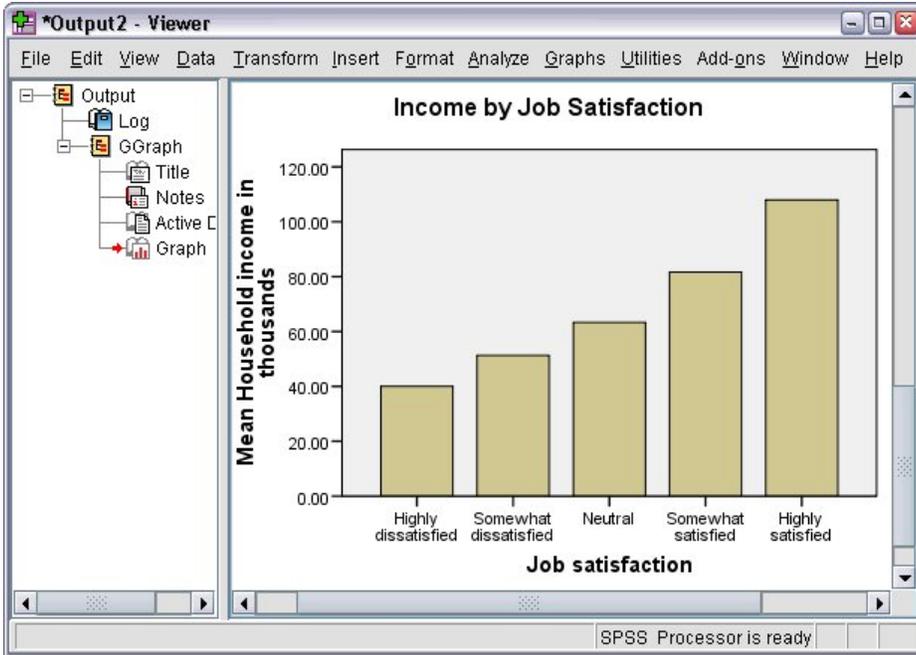


그림 36. 막대도표

이 막대도표에는 직업 만족도가 높은 반응자가 가구소득도 높은 것으로 나타나 있습니다.

## 제 6 장 결과 작업

통계 프로시저를 실행한 결과가 뷰어에 표시됩니다. 프로시저를 실행할 때 선택한 사항에 따라 결과는 통계표, 도표, 그래프, 또는 텍스트 등으로 생성될 수 있습니다. 이 섹션은 파일 *viewertut.spv* 및 *demo.sav* 를 사용합니다. 자세한 정보는 91 페이지의 제 10 장 『표본 파일』 주제를 참조하십시오.

### 뷰어 사용

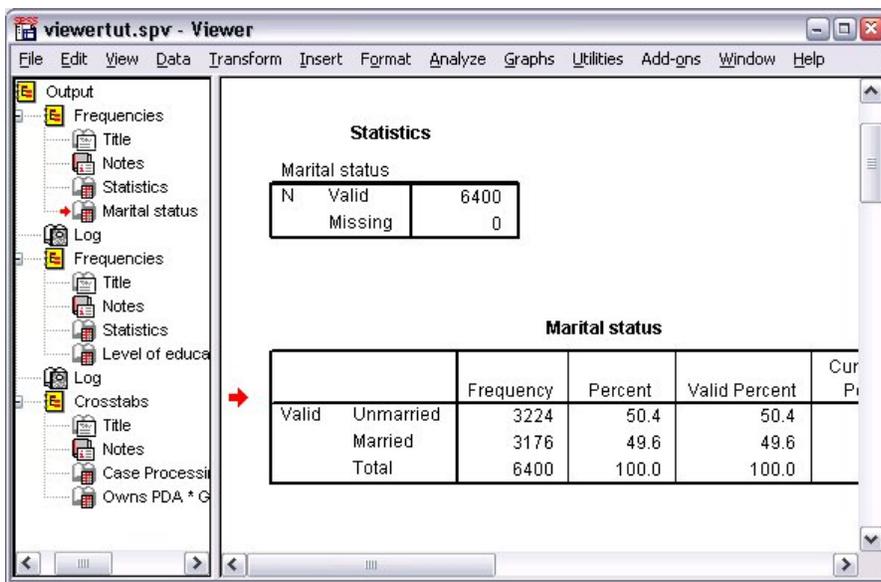


그림 37. 뷰어

뷰어 창은 두 개의 분할창으로 구분됩니다. 개요 분할창에는 뷰어에 저장된 모든 정보의 개요가 표시됩니다. 목차 분할창에는 통계표, 도표 및 텍스트 결과가 표시됩니다.

가로 및 세로 스크롤 막대를 사용하여 창의 내용을 탐색합니다. 보다 쉽게 탐색하려면 개요 분할창에서 항목을 클릭하여 목차 분할창에 항목을 표시합니다.

1. 개요 분할창의 오른쪽 경계선을 클릭한 다음 끌어서 너비를 변경합니다.

개요 분할창에서 항목 옆의 열린 책 모양 아이콘은 현재 목차 분할창에서는 보지 못할 수도 있지만 뷰어에서 볼 수 있는 항목을 나타냅니다.

2. 표가나 도표를 감추려면 개요 분할창에서 책 모양 아이콘을 두 번 클릭합니다.

열린 책 모양 아이콘은 연관된 정보가 현재 감춰져 있음을 나타내는 닫힌 책 모양 아이콘으로 변경됩니다.

3. 감춘 결과를 다시 표시하려면 닫힌 책 모양 아이콘을 두 번 클릭합니다.

또한 특정 통계 프로시저의 모든 결과나 뷰어의 모든 결과를 감출 수 있습니다.

- 특정 프로시저의 결과를 감추려면 해당 프로시저의 왼쪽에서 뿔셈 부호(-)가 있는 상자를 클릭하고 결과를 모두 감추려면 개요 분할창의 맨 위 항목 옆에 있는 상자를 클릭합니다.

개요가 합쳐지면서 이러한 결과가 감춰진 것처럼 보이게 됩니다.

결과가 표시되는 순서를 변경할 수도 있습니다.

- 개요 분할창에서 이동할 항목을 클릭합니다.
- 개요 분할창의 새 위치로 선택한 항목을 끌어다 놓습니다.

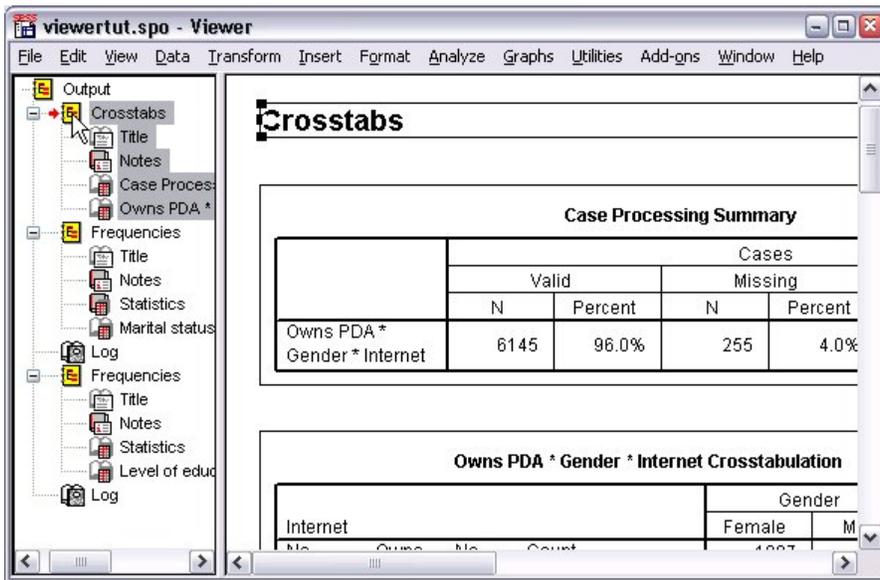


그림 38. 뷰어에서 순서가 변경된 결과

또한 목차 분할창에서 결과 항목을 클릭하여서 끌면 항목을 이동할 수 있습니다.

## 피벗 표 편집기 사용

대부분의 통계 프로시저는 그 결과가 피벗 표로 표시됩니다.

### 결과 정의 액세스

많은 통계 용어가 결과에 표시됩니다. 이러한 용어에 대한 정의를 뷰어에서 직접 액세스할 수 있습니다.

- PDA 소유 \* 성별 \* 인터넷 교차 분석표를 두 번 클릭합니다.
- 기대개수를 마우스 오른쪽 단추로 클릭하고 팝업 메뉴에서 설명을 선택합니다.

팝업 창에 정의가 표시됩니다.

viewertut.spv - Viewer

File Edit View Insert Pivot Format Analyze Graphs Utilities Add-ons Window Help

**Owns PDA \* Gender \* Internet Crosstabulation**

Internet			Gender		Total	
			Female	Male		
No	Owns PDA	No	Count	1897	1962	3859
			Expected Count	1903.4	1955.6	3859.0
		% within Owns PDA	50.3%	49.7%	100.0%	
	Total	Count	2224	2285	4509	
		Expected Count	2224.0	2285.0	4509.0	
Yes	Owns PDA	No	Count	513	512	1025
			Expected Count	512.5	512.5	1025.0
		% within Owns PDA	50.0%	50.0%	100.0%	
	Yes	Count	305	306	611	
		Expected Count	305.5	305.5	611.0	

The number of cases that would be expected in the cell if the row and column variables are statistically independent or unrelated to one another.

그림 39. 팝업 정의

## 피벗 표

생성된 기본값 표에는 사용자가 원하는 것만큼 정보가 명확하게 표시되지 않을 수 있습니다. 피벗 표를 사용해서 행과 열을 전치(표 "이동")하고 표에서 데이터 순서를 조정하며 여러 다른 방법으로 표를 수정할 수 있습니다. 예를 들어, 행과 열을 전치하여 가로로 긴 표를 세로로 길게 변경할 수 있습니다. 표 윤곽을 변경해도 결과에는 영향을 미치지 않습니다. 이러한 방법을 통해 정보를 단지 다른 방식이나 더 적합한 방식으로 표시하는 것입니다.

1. 이것이 이미 활성화되지 않았으면 PDA 소유 \* 성별 \* 인터넷 교차 분석표 표를 두 번 클릭하여 활성화합니다.
2. 피벗 트레이 창이 표시되지 않으면 메뉴에서 다음을 선택합니다.

### 피벗 > 피벗 트레이

피벗 트레이는 열, 행 및 레이어 간에 데이터를 이동하는 방법을 제공합니다.

Internet			Gender		Total
			Female	Male	
Owens PDA	No	Count	1897	1962	3859
		Expected Count	1903.4	1955.6	3859.0
	Yes	% within Owens PDA	49.2%	50.8%	100.0%
		Count	327	323	650
Total					

그림 40. 피벗 트레이

3. 통계 요소를 행 차원에서 성별 아래 열 차원으로 끌어 놓습니다. 표가 바로 재구성되어 변경 사항이 적용됩니다.

피벗 트레이에서 요소의 순서는 표에서 요소의 순서를 반영합니다.

4. 행 차원에서 PDA 소유 요소를 인터넷 요소 앞으로 끌어다 놓아 이 두 행의 순서를 바꿉니다.

			Gender					Total			
			Female			Male					
Internet			Count	Expected Count	% within Owens PDA	Count	Expected Count	% within Owens PDA	Count	Expected Count	% within Owens PDA
Owens PDA	No	No	1897	1903.4	49.2%	1962	1955.6	50.8%	3859	3859.0	100.0%
	Yes	No	327	327.0	100.0%	323	323.0	100.0%	650	1025.0	100.0%
Total	No	Yes	650	611.0	94.0%	650	611.0	94.0%	1300	1222.0	94.0%
	Total		4509	4509.0	100.0%	4509	4509.0	100.0%	9018	1636.0	100.0%

그림 41. 행 바꾸기

## 레이어 작성 및 표시

레이어는 중첩된 정보 범주가 있는 큰 표에 유용합니다. 레이어를 작성하여 표의 모양을 단순화하면 표를 더 쉽게 읽을 수 있습니다.

1. *Gender* 요소를 열 차원에서 레이어 차원으로 끕니다.

		Internet	Count	Expected Count	% within Owns PDA
Owns PDA	No	No	1897	1903.4	49.2%
	Yes	No			
Total	No	Yes			
	Yes	Yes			

그림 42. 레이어 차원의 성별 피벗 아이콘

다른 레이어를 표시하려면 표의 드롭다운 목록에서 범주를 선택합니다.

## 표 편집

사용자 정의 표모양을 작성하지 않은 경우 피벗 표는 표준 형식으로 작성됩니다. 표의 텍스트 형식을 변경할 수 있습니다. 변경할 수 있는 형식은 꼴 이름, 글꼴 크기, 글꼴 스타일(굵게 또는 기울임꼴), 색상 등입니다.

1. 피교육수준 표를 두 번 클릭합니다.
2. 서식 도구 모음이 표시되지 않으면 메뉴에서 다음을 선택합니다.

### 보기 > 도구 모음

3. 제목 텍스트인 피교육수준을 클릭합니다.
4. 도구 모음의 글꼴 크기 드롭다운 목록에서 **12**를 선택합니다.
5. 제목 텍스트의 색상을 변경하려면 텍스트 색상 도구를 클릭하고 새 색상을 선택합니다.



Level of education					
		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	Did not complete high school	1390	21.7	21.7	21.7
	High school degree	1936	30.3	30.3	52.0
	Some college	1360	21.3	21.3	73.2
	College degree	1355	21.2	21.2	94.4
	Post-undergraduate degree	359	5.6	5.6	100.0
	Total	6400	100.0	100.0	

그림 43. 피벗 표에서 형식이 재지정된 제목 텍스트

표의 내용 및 레이블을 편집할 수도 있습니다. 예를 들어, 이 표의 제목을 변경할 수 있습니다.

6. 제목을 두 번 클릭합니다.
7. 새 레이블로 피교육수준을 입력합니다.

참고: 표에서 값을 변경하는 경우 총계와 기타 통계는 다시 계산되지 않습니다.

### 행과 열 감추기

표에 표시된 일부 데이터는 유용하지 않거나 불필요하게 표를 복잡하게 만드는 원인이 될 수 있습니다. 데이터를 손상시키지 않고도 전체 행과 열을 감출 수 있습니다.

1. 이것이 이미 활성화되지 않았으면 피교육수준 표를 두 번 클릭하여 활성화합니다.
2. 유효 퍼센트 열 레이블을 클릭하여 선택합니다.
3. 편집 메뉴나 마우스 오른쪽 단추로 클릭하면 표시되는 팝업 메뉴 메뉴에서 다음을 선택합니다:

선택 > 데이터 및 레이블 셀

4. 보기 메뉴에서 감추기를 선택하거나 마우스 오른쪽 단추로 클릭하면 표시되는 팝업 메뉴에서 범주 감추기를 선택합니다.

열이 감춰지지만 삭제되는 것은 아닙니다.

Education Level					
		Frequency	Percent	Cumulative Percent	
Valid	Did not complete high school	1390	21.7	21.7	
	High school degree	1936	30.3	52.0	
	Some college	1360	21.3	73.2	
	College degree	1355	21.2	94.4	
	Post-undergraduate degree	359	5.6	100.0	
	Total	6400	100.0		

그림 44. 표에서 감춰진 유효 퍼센트 열

열을 다시 표시하려면 다음과 같이 하십시오.

5. 메뉴에서 다음을 선택합니다.

**보기 > 모두 표시**

열과 동일한 방법으로 행을 감추고 표시할 수 있습니다.

### **데이터 표시 형식 바꾸기**

피벗 표에서 데이터 표시 형식을 쉽게 변경할 수 있습니다.

1. 이것이 이미 활성화되지 않았으면 피교육수준 표를 두 번 클릭하여 활성화합니다.
2. 퍼센트 열 레이블을 클릭하여 선택합니다.
3. 편집 메뉴나 마우스 오른쪽 단추로 클릭하면 표시되는 팝업 메뉴 메뉴에서 다음을 선택합니다:

**선택 > 데이터 셀**

4. 형식 메뉴 또는 마우스 오른쪽 단추로 클릭하면 표시되는 팝업 메뉴 메뉴에서 **셀 특성**을 선택합니다.
5. **형식 값** 탭을 클릭합니다.
6. 소수점이하자리 필드에 0을 입력하여 이 열의 모든 소수점을 감춥니다.

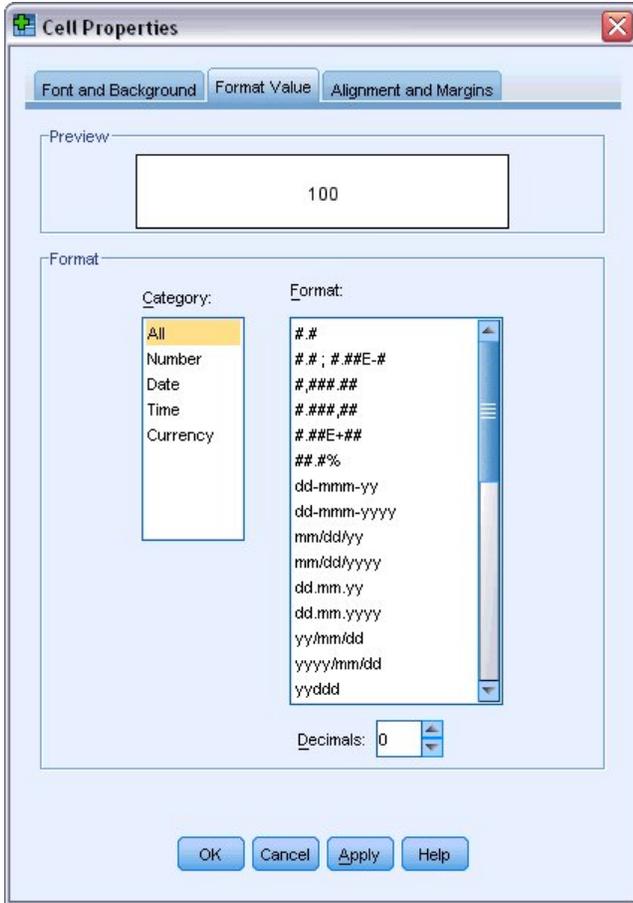


그림 45. 셀 특성, 형식 값 탭

이 대화 상자에서 데이터 유형과 형식도 변경할 수 있습니다.

7. 범주 목록에서 원하는 유형을 선택한 다음 형식 목록에서 해당 유형의 형식을 선택합니다.
8. **확인** 또는 **적용**을 클릭하여 변경사항을 적용합니다.

### Education Level

		Frequency	Percent	Cumulative Percent
Valid	Did not complete high school	1390	22	21.7
	High school degree	1936	30	52.0
	Some college	1360	21	73.2
	College degree	1355	21	94.4
	Post-undergraduate degree	359	6	100.0
	Total	6400	100	

그림 46. 퍼센트 열에서 감추어진 소수점이하자리

이제 퍼센트 열에서 소수점이하자리가 감춰집니다.

## 표모양

표의 형식은 분명하고 간결하며 의미 있는 결과를 제공하는 데 중요한 부분입니다. 표를 읽기 어려우면 그 안에 들어 있는 정보를 쉽게 이해하지 못할 수 있습니다.

### 사전 정의된 형식 사용

1. 결혼 상태 표를 두 번 클릭합니다.
2. 메뉴에서 다음을 선택합니다.

형식 > 표모양...

표모양 대화 상자는 사전 정의된 다양한 스타일을 나열합니다. 목록에서 스타일을 선택하여 오른쪽의 보기 창에서 미리보기를 확인합니다.

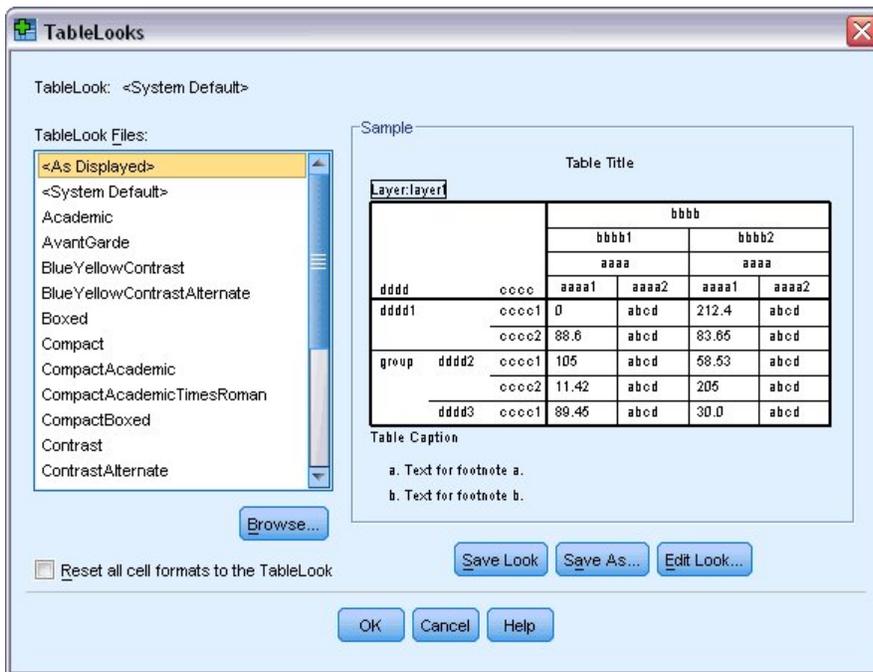


그림 47. 표모양 대화 상자

스타일을 그대로 사용하거나 사용자의 요구에 적합하게 기존 스타일을 편집할 수 있습니다.

3. 기존 스타일을 사용하려면 스타일을 선택하고 확인을 클릭합니다.

### 표모양 스타일 사용자 정의

사용자의 특정 요구에 맞게 형식을 사용자 정의할 수 있습니다. 배경 색상에서 경계선 스타일에 이르기까지 표의 거의 모든 형식을 사용자 정의할 수 있습니다.

1. 결혼 상태 표를 두 번 클릭합니다.
2. 메뉴에서 다음을 선택합니다.

형식 > 표모양...

- 원하는 형식과 가장 비슷한 스타일을 선택하고 **모양 편집**을 클릭하십시오.
- 셀 형식** 탭을 클릭하여 서식 지정 옵션을 표시합니다.

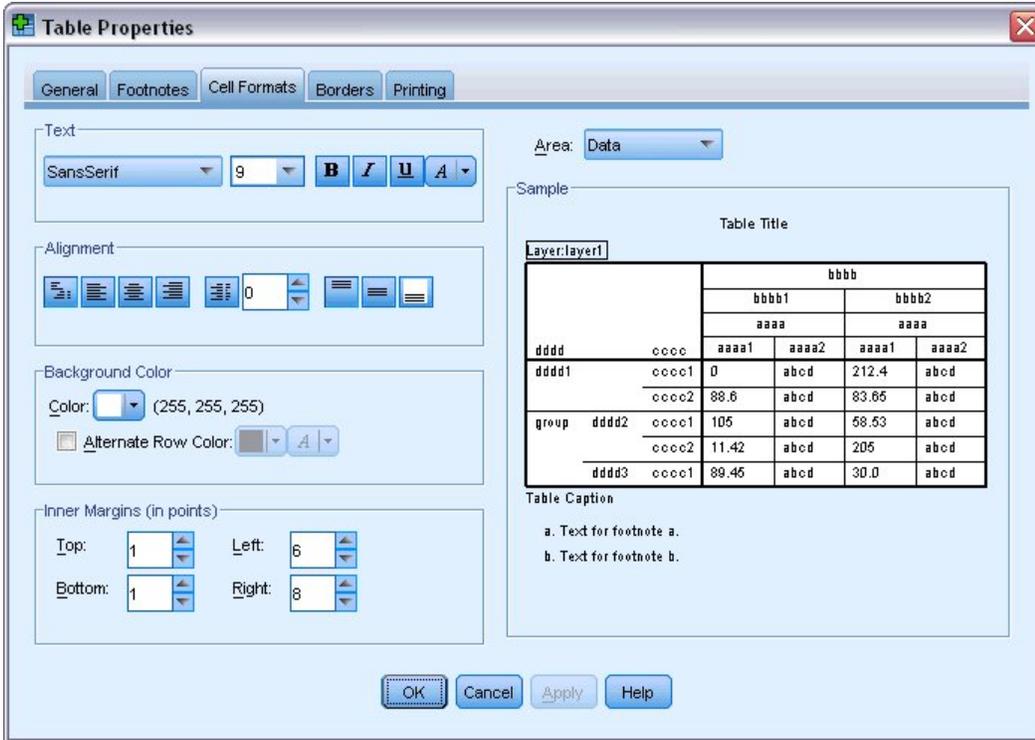


그림 48. 표 특성 대화 상자

서식 지정 옵션에는 글꼴 이름, 글꼴 크기, 스타일 및 색상이 포함됩니다. 추가 옵션으로는 정렬, 텍스트 및 배경 색상, 여백 크기 등이 있습니다.

오른쪽의 보기 창에서는 서식 변경 내용이 표에 적용된 결과를 미리 보여 줍니다. 표의 각 영역에 다른 서식 스타일을 지정할 수 있습니다. 예를 들어, 제목과 데이터의 스타일을 다르게 설정하려는 경우가 있습니다. 편집할 표 영역을 선택하려면 영역 드롭다운 목록에서 이름으로 영역을 선택하거나 보기 창에서 변경하려는 영역을 누를 수 있습니다.

- 영역 드롭다운 목록에서 **데이터**를 선택합니다.
- 배경 색상 드롭다운 팔레트에서 새 색상을 선택합니다.
- 그런 다음 다음 번 텍스트 색상을 선택합니다.

보기 창에 새 스타일이 표시됩니다.

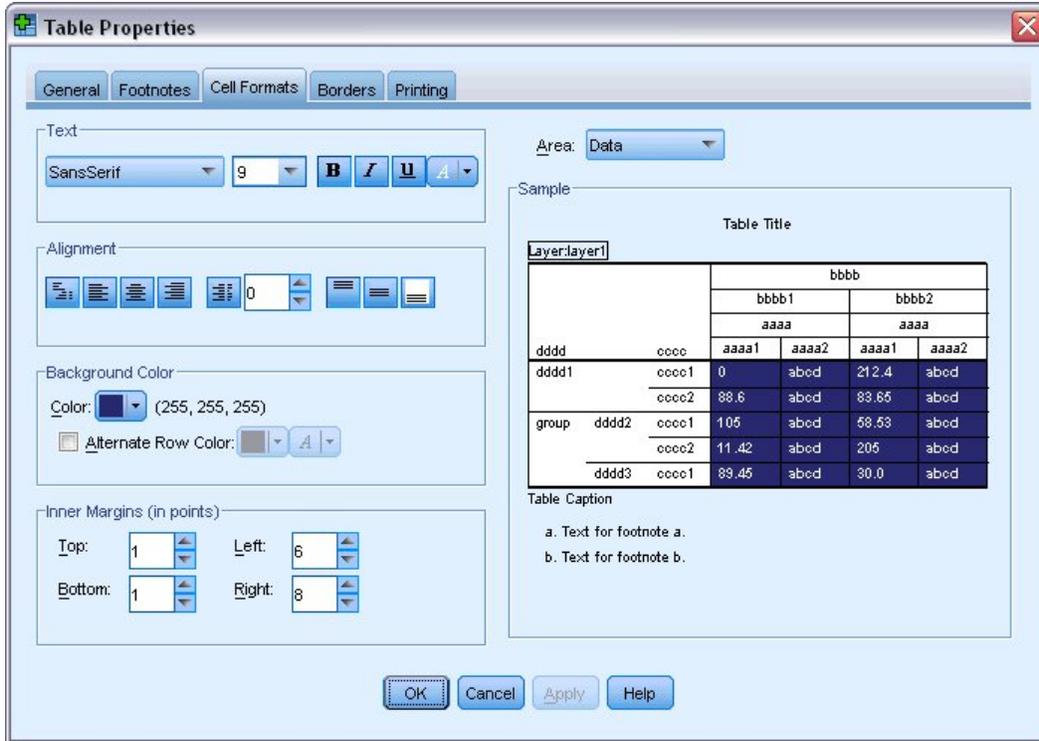


그림 49. 표 셀 형식 변경

8. 확인을 클릭하여 표모양 대화 상자로 돌아갑니다.

나중에 표에 쉽게 적용할 수 있도록 새 스타일을 저장할 수 있습니다.

9. 다른 이름으로 저장을 클릭합니다.

10. 대상 디렉토리로 이동하여 새 스타일의 이름을 파일 이름 입력란에 입력합니다.

11. 저장을 클릭합니다.

12. 확인을 클릭하여 변경 사항을 적용하고 뷰어로 돌아갑니다.

지정된 사용자 정의 형식이 표에 포함됩니다.

The screenshot shows a window titled 'viewertut.spv' with a menu bar (File, Edit, View, Insert, Pivot, Format, Analyze, Graphs, Utilities, Add-ons, Window, Help). The main content is a pivot table titled 'Marital status' with the following data:

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	Unmarried	3224	50.4	50.4	50.4
	Married	3176	49.6	49.6	100.0
	Total	6400	100.0	100.0	

그림 50. 사용자 정의 표모양

## 기본값 표 형식 바꾸기

표를 작성한 후 표 형식을 변경할 수 있지만 표를 작성할 때마다 형식을 변경할 필요가 없도록 기본값 표모양을 변경하는 것이 더 효율적일 수 있습니다.

피벗 표에 대한 기본값 표모양 스타일을 변경하려면 메뉴에서 다음을 선택합니다.

편집 > 옵션...

1. 옵션 대화 상자에서 **피벗 표** 탭을 클릭합니다.

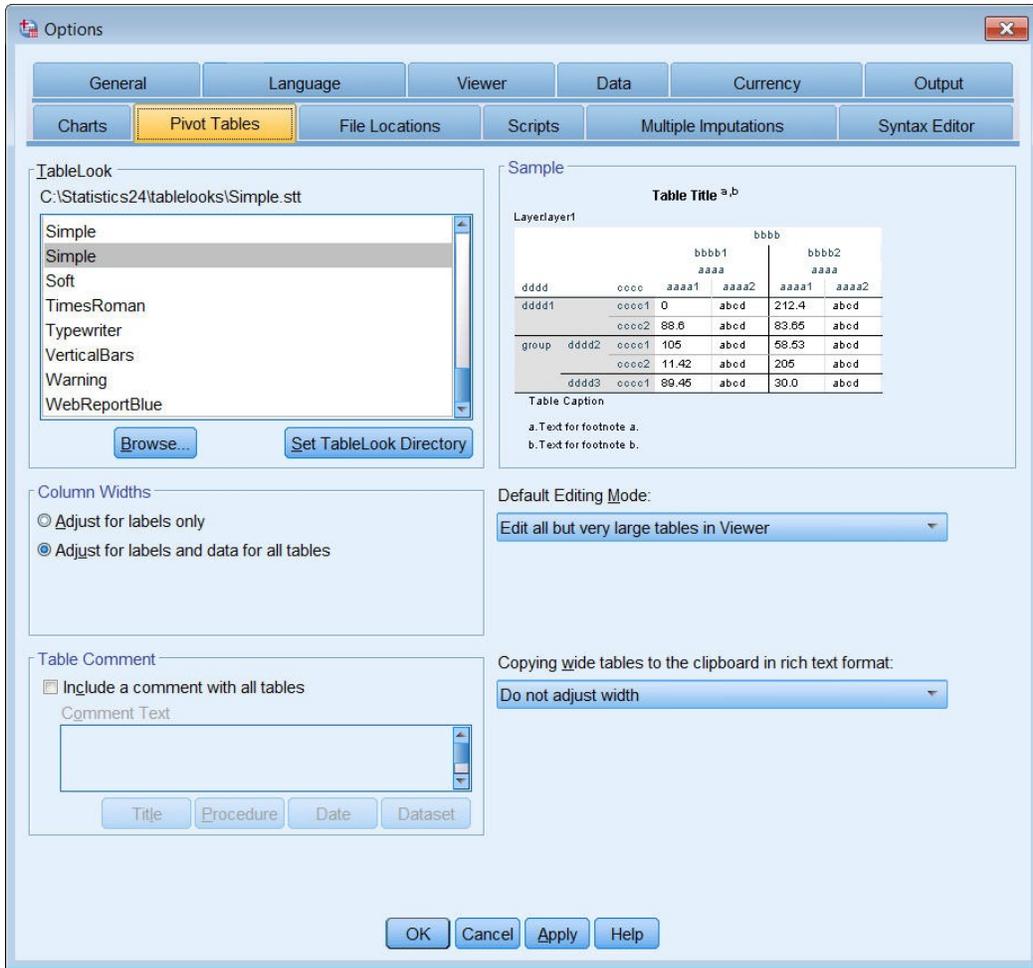


그림 51. 옵션 대화 상자

2. 모든 새 표에 사용할 표모양 스타일을 선택합니다.

오른쪽의 보기 창에 각 표모양의 미리보기가 표시됩니다.

3. 확인을 클릭하여 설정을 저장하고 대화 상자를 닫습니다.

기본값 표모양을 변경한 후 작성하는 모든 표는 자동으로 새 형식 규칙을 따릅니다.

### 초기 출력 설정 사용자 정의

초기 출력 설정에는 뷰어에서 개체 맞춤, 개체를 기본적으로 보이기 또는 감추기로 설정할지 여부 및 뷰어 창 너비가 포함됩니다. 이러한 설정을 변경하려면 다음과 같이 하십시오.

1. 메뉴에서 다음을 선택합니다.

편집 > 옵션...

2. 뷰어 탭을 클릭합니다.

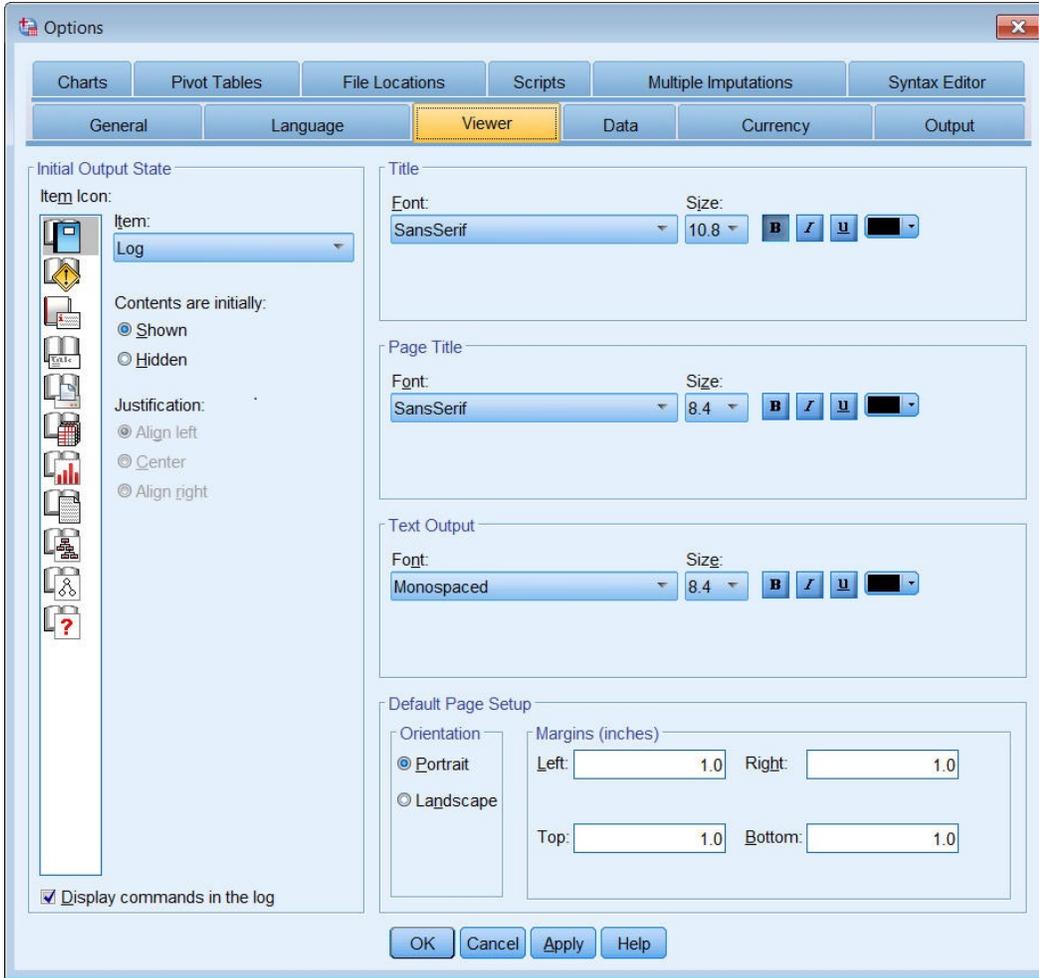


그림 52. 뷰어 옵션 설정

설정은 개체별로 적용됩니다. 예를 들어, 표가 표시되는 방법은 변경하지 않고 도표가 표시되는 방법을 사용자 정의할 수 있습니다. 사용자 정의할 개체를 선택하고 원하는 사항을 변경합니다.

3. 제목 아이콘을 클릭하여 해당 설정을 표시합니다.
4. 가운데 맞춤을 클릭하여 뷰어의 가운데(수평)에 모든 제목을 표시합니다.

또한 결과를 복잡하게 만들 수 있는 로그 및 경고 메시지 등의 요소를 감출 수 있습니다. 아이콘을 두 번 클릭하면 해당 개체의 표시 특성이 자동으로 변경됩니다.

5. 경고 아이콘을 두 번 클릭하여 결과에서 경고 메시지를 감춥니다.
6. 확인을 클릭하여 변경 사항을 저장하고 대화 상자를 닫습니다.

## 변수와 값 레이블 표시

대부분의 경우 변수 및 변수값에 대한 레이블을 표시하는 것이 변수 이름이나 실제 데이터 값을 표시하는 것보다 효과적입니다. 하지만 이름과 레이블을 모두 표시하려는 경우가 있을 수 있습니다.

1. 메뉴에서 다음을 선택합니다.

편집 > 옵션...

2. 결과 레이블 탭을 클릭합니다.

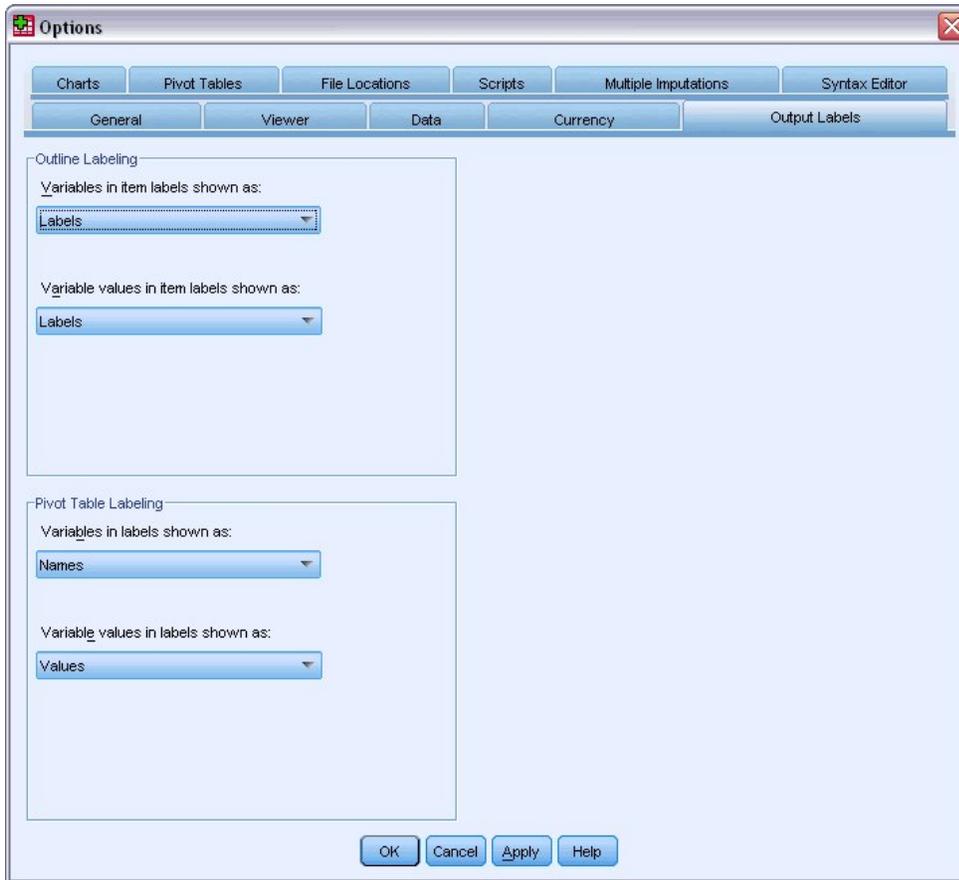


그림 53. 피벗 표 레이블 붙이기 설정

개요 분할창 및 목차 분할창에 대해 다른 설정을 지정할 수 있습니다. 예를 들어, 개요 분할창에 레이블을 표시하고 목차 분할창에 변수 이름과 데이터 값을 표시하려면 다음과 같이 하십시오.

3. 피벗 표 레이블 붙이기에서 변수의 레이블 표시는 다음과 같이 나타남 드롭다운 목록에서 이름을 선택하여 레이블 대신 변수 이름을 표시합니다.
4. 그런 다음 변수값의 레이블 표시는 다음과 같이 나타남 드롭다운 목록에서 값을 선택하여 레이블 대신에 데이터 값을 표시합니다.

이후에 세션에서 생성된 표는 이러한 변경사항이 반영됩니다.

marital

	Frequency	Percent	Valid Percent	Cumulative Percent
Valid 0	3224	50.4	50.4	50.4
1	3176	49.6	49.6	100.0
Total	6400	100.0	100.0	

그림 54. 표시된 변수 이름과 값

## 다른 애플리케이션에서 결과 사용

결과를 여러 애플리케이션에서 사용할 수 있습니다. 예를 들어, 프리젠테이션이나 보고서에 표가나 도표를 포함할 수 있습니다.

다음 예제는 Microsoft Word에 해당하는 것이지만 다른 워드 프로세스 애플리케이션에서도 유사한 방식으로 사용할 수 있습니다.

### Word 표로 결과 붙여넣기

피벗 표를 원시 Word 표로 Word에 붙여넣을 수 있습니다. 글꼴 크기 및 색상과 같은 모든 표 속성이 유지됩니다. 표를 Word 표 형식으로 붙여넣기 때문에 다른 모든 표와 마찬가지로 Word에서 편집할 수 있습니다.

1. 뷰어에서 표를 클릭하여 선택합니다.
2. 메뉴에서 다음을 선택합니다.

**편집 > 복사**

3. 워드 프로세스 애플리케이션을 엽니다.
4. 워드 프로세서의 메뉴에서 다음을 선택합니다.

**편집 > 선택하여 붙여넣기...**

5. 선택하여 붙여넣기 대화 상자에서 **서식있는 텍스트(RTF)**를 선택합니다.
6. **확인**을 클릭하여 결과를 현재 문서에 붙여넣습니다.

이제 문서에 표가 표시됩니다. 필요에 따라 적합하게 사용자 정의 형식을 적용하고 데이터를 편집하며 표 크기를 변경할 수 있습니다.

### 텍스트로 결과 붙여넣기

피벗 표를 일반 텍스트로 다른 애플리케이션에 복사할 수 있습니다. 이 방법에서는 서식 스타일이 유지되지 않지만 대상 애플리케이션에 표 데이터를 붙여넣은 후 이를 편집할 수 있습니다.

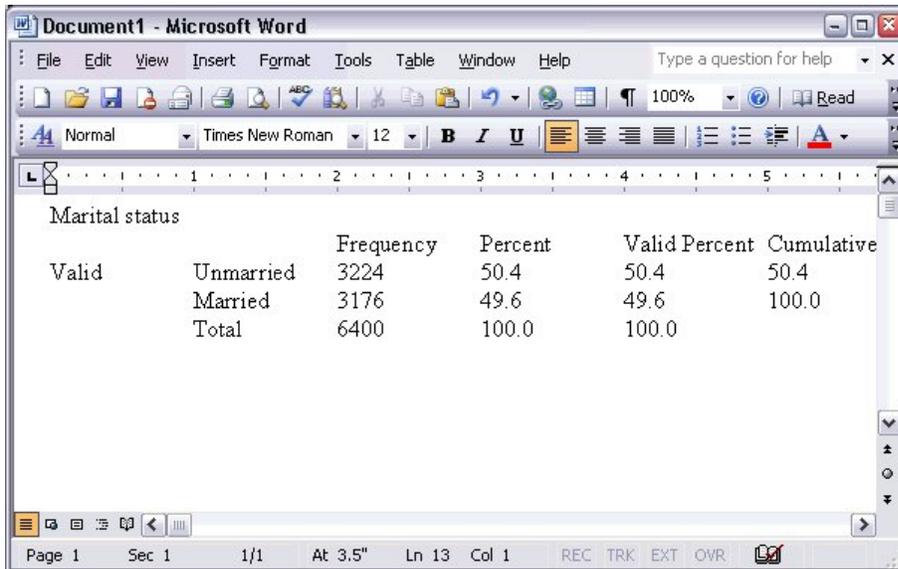
1. 뷰어에서 표를 클릭하여 선택합니다.
2. 메뉴에서 다음을 선택합니다.

### 편집 > 복사

3. 워드 프로세스 애플리케이션을 엽니다.
4. 워드 프로세서의 메뉴에서 다음을 선택합니다.

### 편집 > 선택하여 붙여넣기...

5. 선택하여 붙여넣기 대화 상자에서 서식없는 텍스트를 선택합니다.
6. 확인을 클릭하여 결과를 현재 문서에 붙여넣습니다.



The screenshot shows a Microsoft Word window titled 'Document1 - Microsoft Word'. The main content area displays a pivot table with the following data:

Marital status		Frequency	Percent	Valid Percent	Cumulative
Valid	Unmarried	3224	50.4	50.4	50.4
	Married	3176	49.6	49.6	100.0
Total		6400	100.0	100.0	

그림 55. Word에 표시된 피벗 표

표의 각 열은 탭으로 구분됩니다. 워드 프로세스 애플리케이션에서 탭 간격을 조정하여 열 너비를 변경할 수 있습니다.

## Microsoft Word, PowerPoint 및 Excel 파일로 결과 내보내기

Microsoft Word, PowerPoint 또는 Excel 파일로 결과를 내보낼 수 있습니다. 선택한 항목 또는 뷰어의 모든 항목을 내보낼 수 있습니다. 이 섹션은 파일 *msouttut.spv* 및 *demo.sav*를 사용합니다. 자세한 정보는 91 페이지의 제 10 장 『표본 파일』 주제를 참조하십시오.

참고: PowerPoint로 내보내기는 Windows 운영체제에서만 사용할 수 있으며 Student 버전에 대해서는 사용할 수 없습니다.

뷰어의 개요 분할창에서 내보낼 특정 항목을 선택하거나 모든 항목 또는 볼 수 있는 모든 항목을 내보낼 수 있습니다.

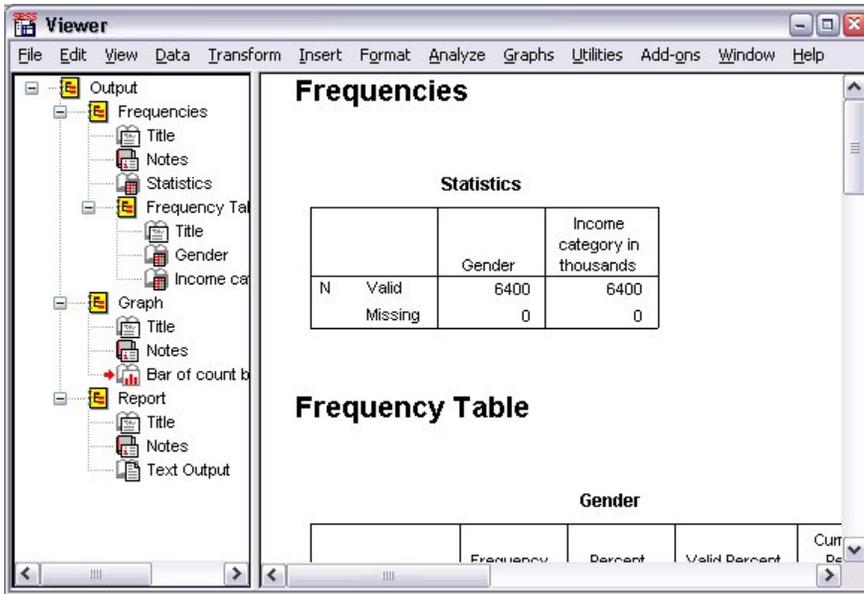


그림 56. 뷰어

1. 뷰어 메뉴에서 다음을 선택합니다.

**파일 > 내보내기...**

뷰어에서 모든 개체를 내보내지 않고 표시된 개체(개요 분할창의 열린 책 모양)만 내보내거나 개요 분할창에서 선택한 개체만 내보내도록 선택할 수 있습니다. 개요 분할창에서 항목을 선택하지 않은 경우 선택한 개체를 내보내는 옵션이 표시되지 않습니다.

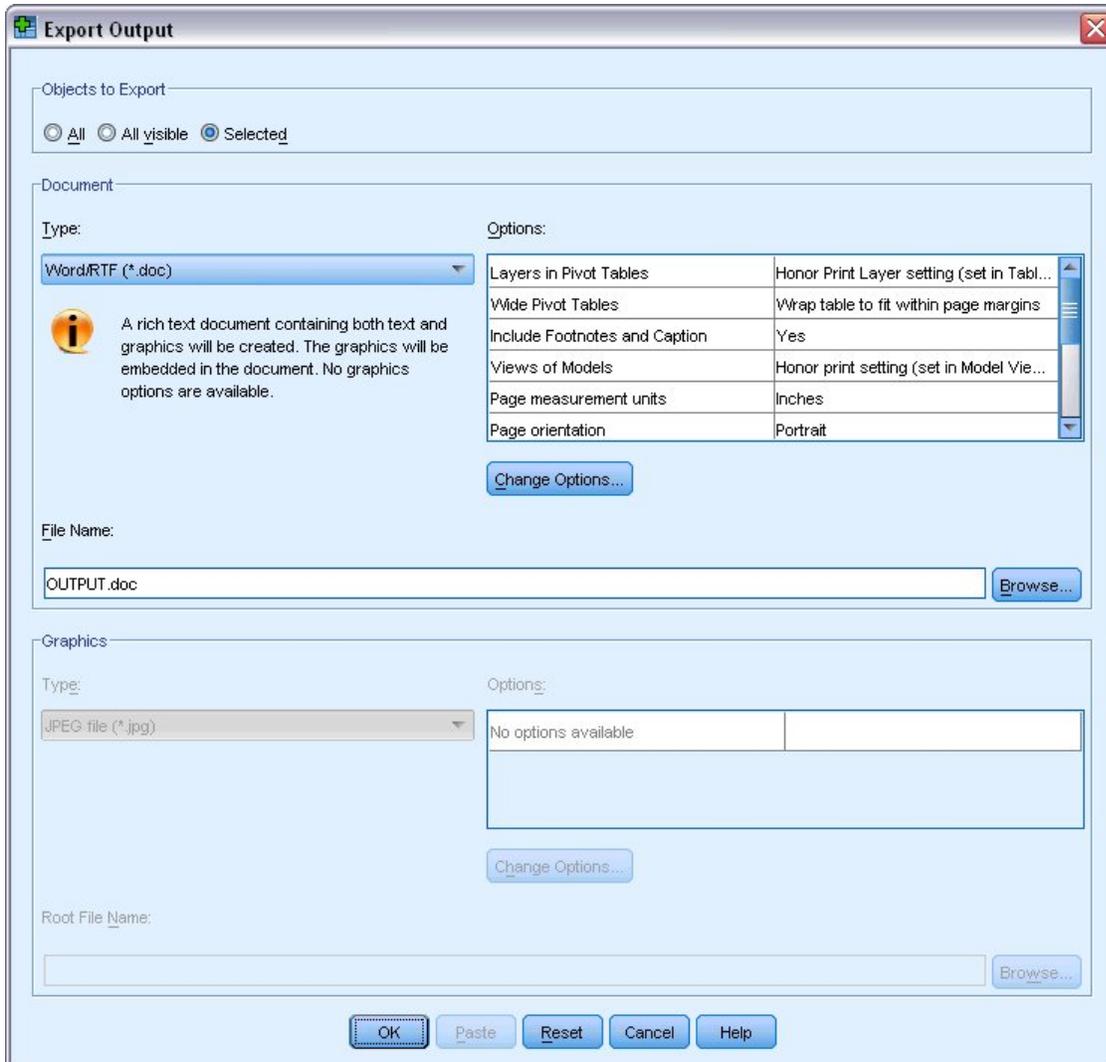


그림 57. 내보내기 결과 대화 상자

2. 내보내기 그룹 개체에서 모두를 선택합니다.
3. 유형 드롭다운 목록에서 **Word/RTF 파일 (\*.doc)**을 선택합니다.
4. **확인**을 클릭하여 Word 파일을 생성합니다.

Word에서 결과 파일을 열면 결과를 내보낸 방법을 확인할 수 있습니다. 모든 개체를 내보내도록 선택했으므로 표시된 개체가 아닌 노트가 Word에 표시됩니다.

피벗 표는 글꼴, 색상, 경계선 등을 비롯한 원래 피벗 표의 모든 형식을 유지하는 Word 표가 됩니다.

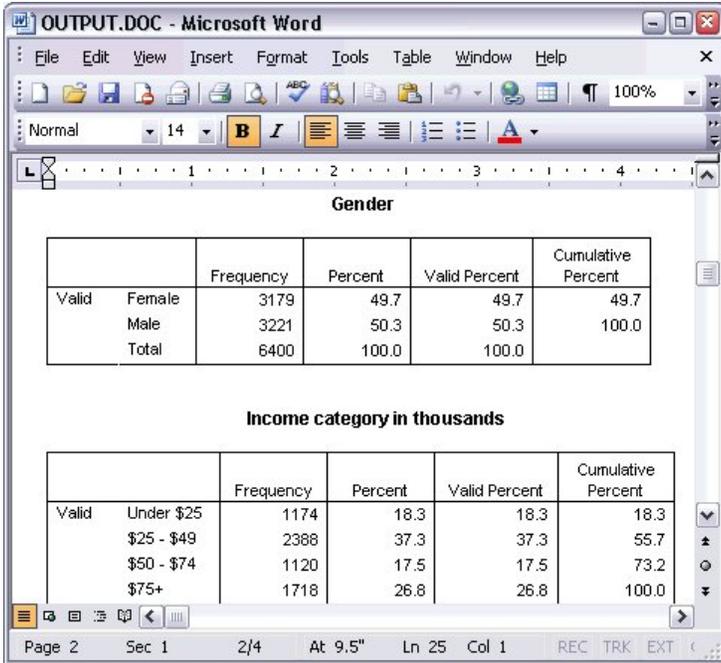


그림 58. Word에서의 피벗 표

도표는 그래픽 이미지로 Word 문서에 포함됩니다.

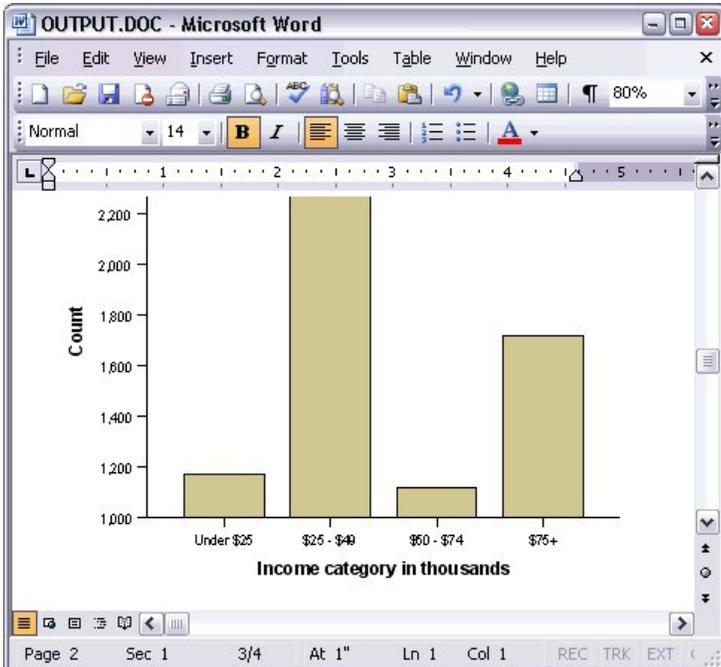


그림 59. Word에서의 도표

텍스트 결과는 뷰어에서 텍스트 개체에 사용되는 글꼴과 동일한 글꼴로 표시됩니다. 적절한 정렬을 위해 텍스트 결과는 고정 피치(단일자간) 글꼴을 사용해야 합니다.

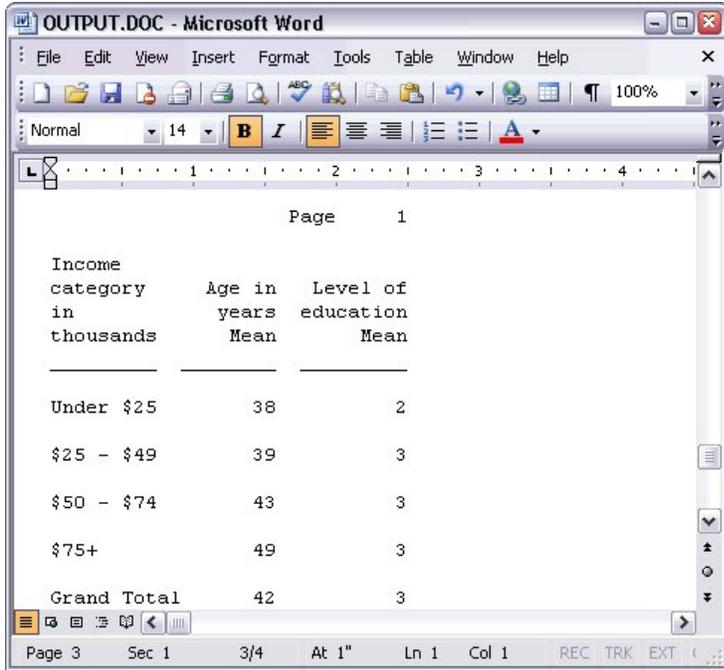


그림 60. Word에서의 텍스트 결과

PowerPoint 파일로 내보내는 경우 내보낸 각 항목은 별도의 슬라이드에 배치됩니다. PowerPoint로 내보낸 피벗 표는 글꼴, 색상, 경계선 등을 포함하여 원래 피벗 표의 모든 서식을 유지하는 Word 표가 됩니다.

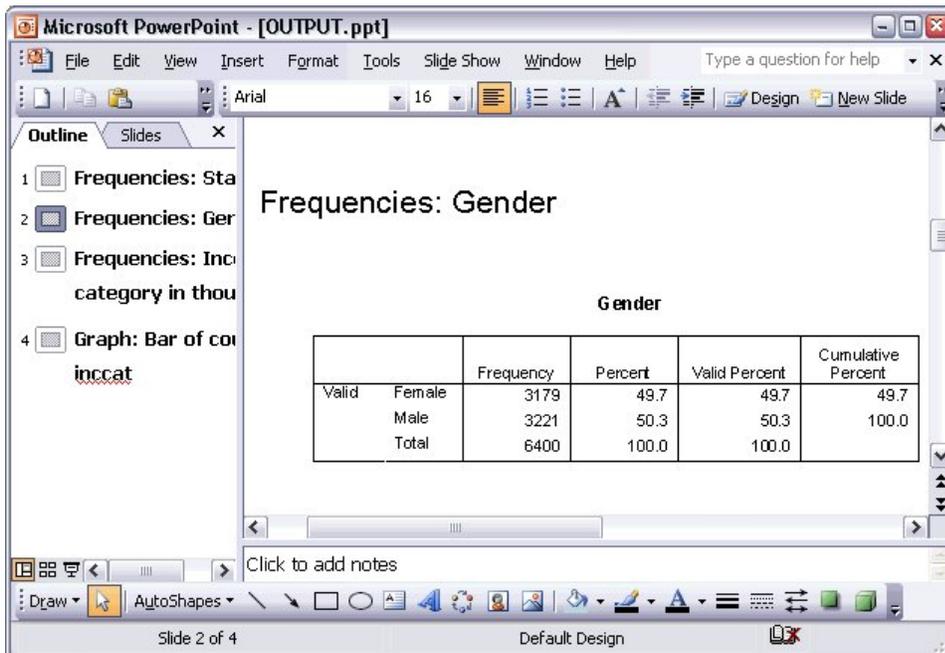


그림 61. PowerPoint에서의 피벗 표

PowerPoint로 내보낼 도표는 PowerPoint 파일에 포함됩니다.

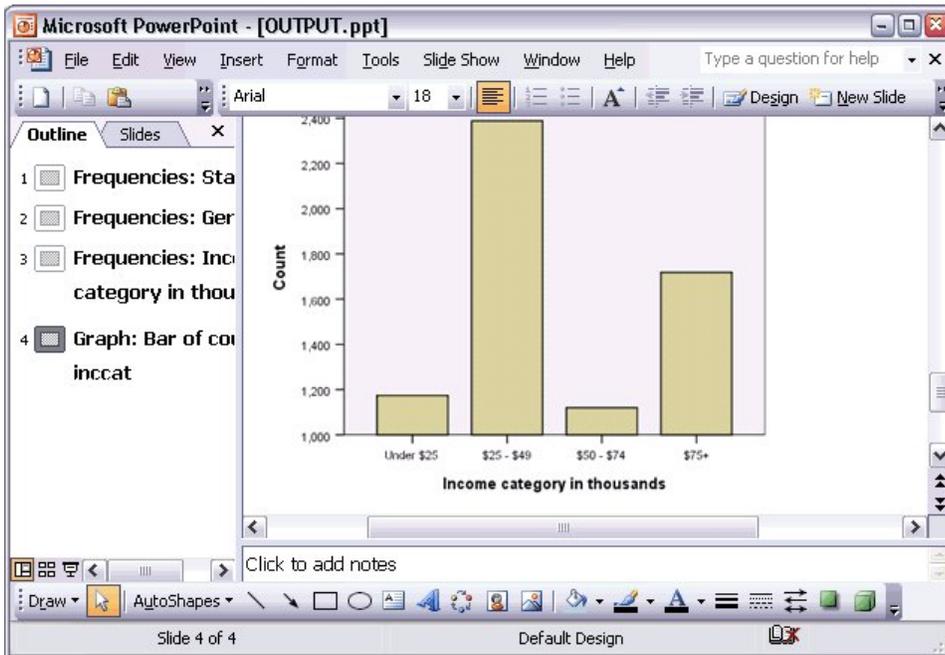


그림 62. PowerPoint에서의 도표

참고: PowerPoint로 내보내기는 Windows 운영체제에서만 사용할 수 있으며 Student 버전에 대해서는 사용할 수 없습니다.

Excel 파일로 내보낼 경우 결과가 다르게 내보내집니다.

Microsoft Excel - OUTPUT.xls

		Frequency	Percent	Valid Percent	Cumulative Percent
<b>Gender</b>					
Valid	Female	3,179	49.7	49.7	49.7
	Male	3,221	50.3	50.3	100.0
	Total	6,400	100.0	100.0	
<b>Income category in thousands</b>					
Valid	Under \$25	1,174	18.3	18.3	18.3
	\$25 - \$49	2,388	37.3	37.3	55.7
	\$50 - \$74	1,120	17.5	17.5	73.2
	\$75+	1,718	26.8	26.8	100.0
	Total	6,400	100.0	100.0	

그림 63. Excel에서의 Output.xls

피벗 표 행, 열 및 셀은 Excel 행, 열 및 셀이 됩니다.

텍스트 결과의 각 줄은 Excel 파일에서 하나의 행이고, 이 행은 단일 셀에 포함된 줄의 전체 내용을 포함합니다.

Microsoft Excel - OUTPUT.xls

Income category in thousands	Age in years Mean	Level of education Mean
Under \$25	38	2
\$25 - \$49	39	3
\$50 - \$74	43	3
\$75+	49	3
Grand Total	42	3

그림 64. Excel에서의 텍스트 결과

## PDF로 결과 내보내기

뷰어에서 PDF(Portable Document Format) 파일로 모든 항목 또는 선택된 항목을 내보낼 수 있습니다.

1. PDF로 내보내려는 결과를 포함하는 뷰어 창의 메뉴에서 다음을 선택합니다:

파일 > 내보내기...

2. 내보내기 결과 대화 상자의 내보내기 형식 파일 유형 드롭다운 목록에서 **Portable Document Format**을 선택합니다.

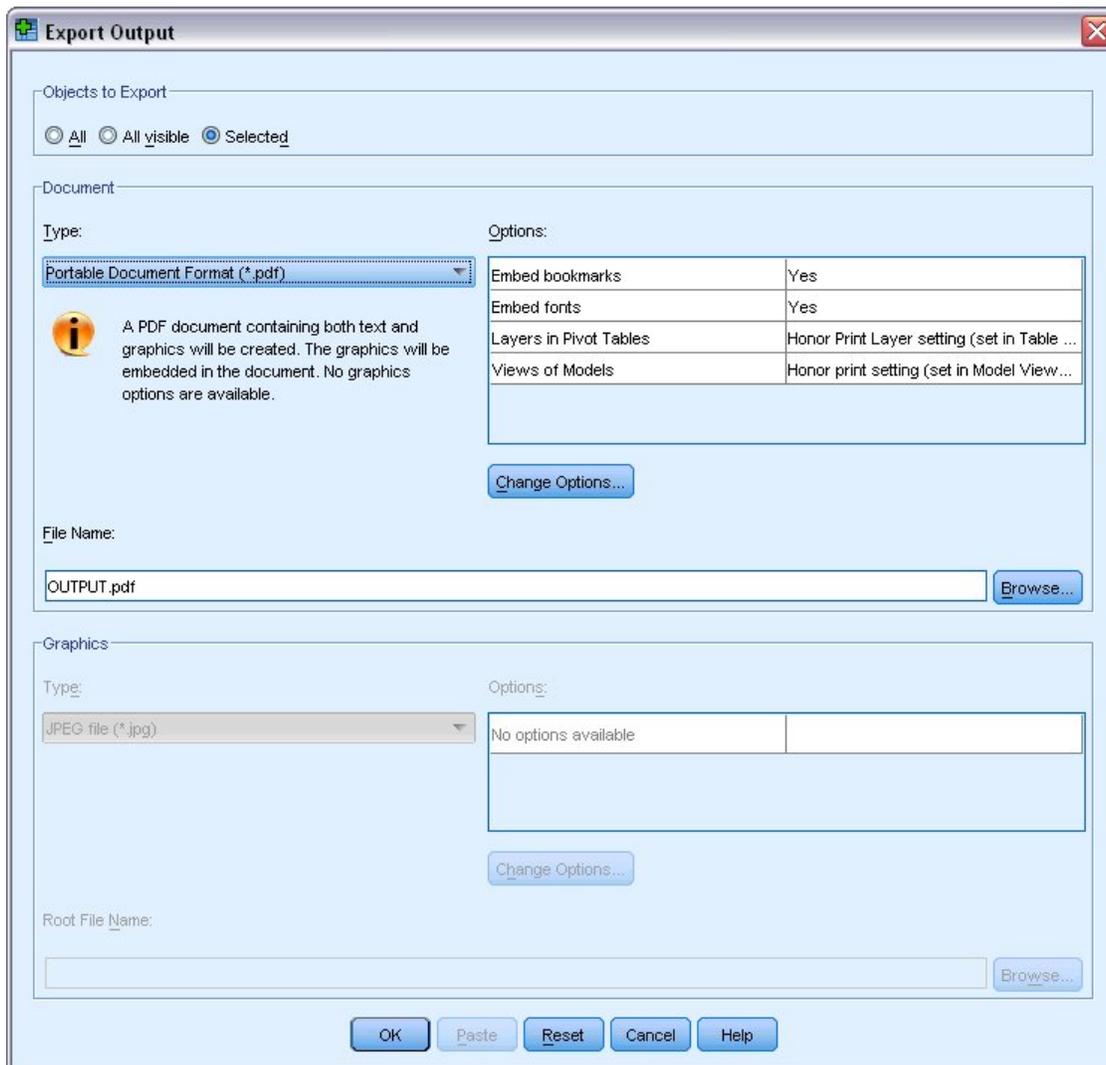


그림 65. 내보내기 결과 대화 상자

- 뷰어 문서의 개요 분할창은 쉽게 탐색할 수 있도록 PDF 파일의 책갈피로 전환됩니다.
- PDF 문서의 쪽 크기, 방향, 여백, 쪽 머리글과 꼬리말의 내용 및 출력, 인쇄된 문자 크기는 뷰어 창의 파일 메뉴에서 쪽 설정을 선택하여 액세스할 수 있는 쪽 설정 옵션으로 제어합니다.

- PDF 문서의 해상도(DPI)는 기본값 프린터 또는 현재 선택된 프린터의 현재 해상도 설정입니다. 이는 쪽 설정을 사용하여 변경할 수 있습니다. 최대 해상도는 1200DPI입니다. 프린터 설정이 이보다 더 높더라도 PDF 문서는 1200DPI로 설정됩니다. 참고: 문서의 해상도가 높아도 프린터의 해상도가 낮으면 인쇄 결과물의 품질이 낮을 수 있습니다.

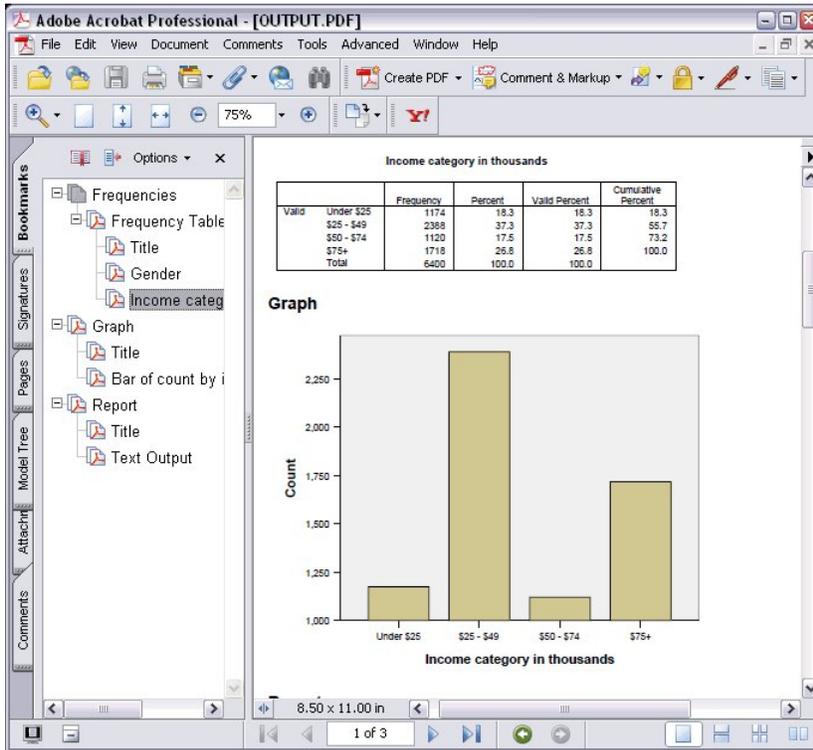


그림 66. 책갈피가 있는 PDF 파일

## HTML로 결과 내보내기

결과를 HTML(Hypertext Markup Language)로도 내보낼 수 있습니다. HTML로 저장할 경우 그래픽이 아닌 모든 결과를 단일 HTML 파일로 내보냅니다.

The screenshot shows two tables from an SPSS output file. The first table is titled 'Gender' and the second is 'Income category in thousands'. Both tables show Frequency, Percent, Valid Percent, and Cumulative Percent for each category.

	Frequency	Percent	Valid Percent	Cumulative Percent
Female	3179	49.7	49.7	49.7
Valid Male	3221	50.3	50.3	100.0
Total	6400	100.0	100.0	

	Frequency	Percent	Valid Percent	Cumulative Percent
Under \$25	1174	18.3	18.3	18.3
\$25 - \$49	2388	37.3	37.3	55.7
Valid \$50 - \$74	1120	17.5	17.5	73.2
\$75+	1718	26.8	26.8	100.0
Total	6400	100.0	100.0	

그림 67. 웹 브라우저에서의 *Output.htm*

HTML로 내보내는 경우 도표도 함께 내보낼 수 있지만 단일 파일로 내보낼 수는 없습니다.

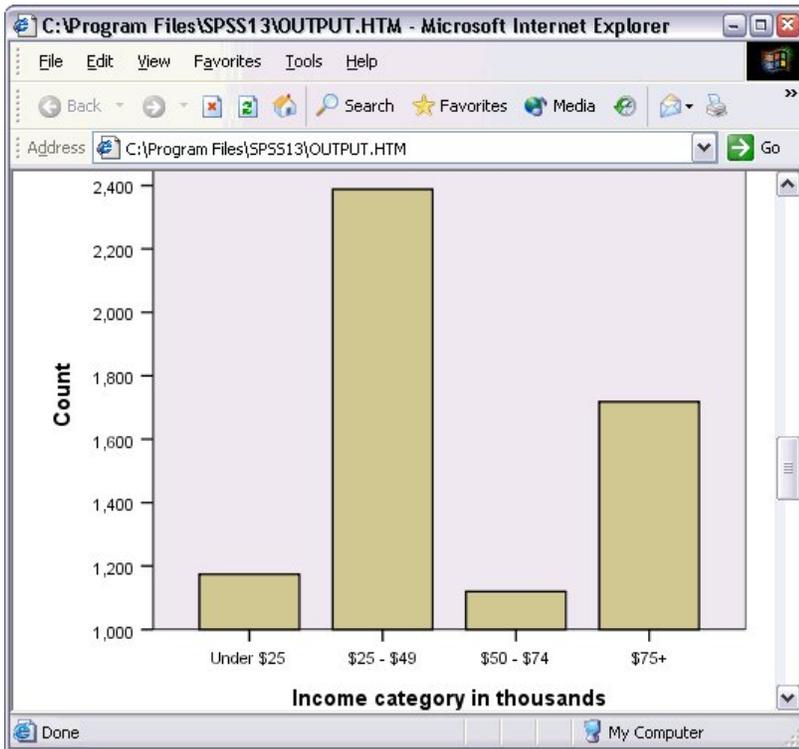


그림 68. HTML에서의 도표

각 도표는 사용자가 지정한 형식의 파일로 저장되며 이러한 그래픽 파일에 대한 참조는 HTML에 저장됩니다. 또한 별도의 그래픽 파일에 모든 도표 또는 선택한 도표를 내보낼지 선택할 수 있습니다.

---

## 제 7 장 구문 사용

강력한 명령을 사용하여 수많은 공통 작업을 저장하고 자동화할 수 있습니다. 메뉴와 대화 상자에 없는 일부 기능도 제공합니다. 대부분의 명령은 메뉴와 대화 상자에서 액세스할 수 있습니다. 그러나 일부 명령과 옵션은 명령을 통해서만 사용할 수 있습니다. 또 명령을 사용하면 명령 구문 파일에 작업을 저장하여 나중에 분석을 반복할 수 있습니다.

명령 구문 파일은 IBM SPSS Statistics 명령이 들어 있는 텍스트 파일에 불과합니다. 구문 창을 열고 명령을 직접 입력할 수 있지만 대화 상자에서 전체 또는 일부 작업을 수행하도록 하는 것이 보다 쉬운 방법입니다.

이 장의 예제에서는 *demo.sav* 데이터 파일을 사용합니다. 자세한 정보는 91 페이지의 제 10 장 『표본 파일』 주제를 참조하십시오.

참고: Student 버전에서는 명령 구문을 사용할 수 없습니다.

---

### 구문 붙여넣기

구문을 생성하려면 대부분의 대화 상자에 있는 구문 단추를 누르기만 하면 됩니다.

1. 데이터 파일 *demo.sav*를 엽니다. 자세한 정보는 91 페이지의 제 10 장 『표본 파일』 주제를 참조하십시오.
2. 메뉴에서 다음을 선택합니다.  
  
    **분석 > 기술통계량 > 빈도...**
3. 결혼 상태 [결혼]을 선택하고 변수 목록으로 이동합니다.
4. 도표를 클릭합니다.
5. 도표 대화 상자에서 **막대도표**를 선택합니다.
6. 도표화 값에서 **퍼센트**를 선택합니다.
7. **계속**을 클릭합니다. 대화 상자에서 선택한 내용에 따라 작성된 구문을 구문 편집기로 복사하려면 **붙여넣기**를 클릭합니다.

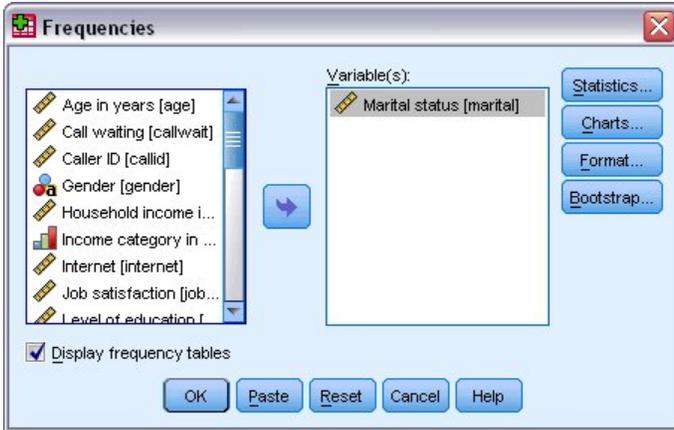


그림 69. 빈도분석 대화 상자

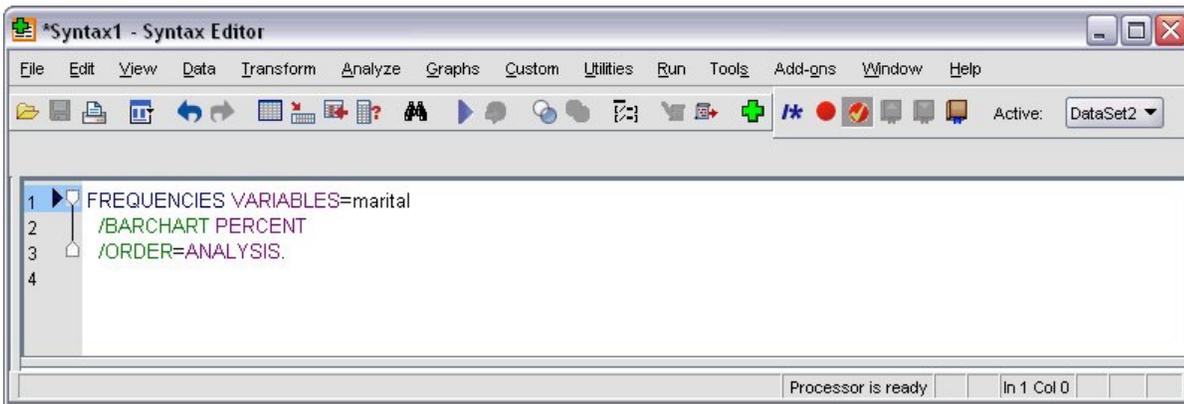


그림 70. 빈도분석 구문

8. 현재 표시된 구문을 실행하려면 메뉴에서 다음을 선택합니다.

실행 > 선택

## 구문 편집

구문 창에서 구문을 편집할 수 있습니다. 예를 들어, 퍼센트 대신 빈도분석이 표시되도록 /BARCHART 부명령문을 변경할 수 있습니다. (부명령문에는 슬래시가 표시됩니다.) 표시 빈도를 위한 키워드를 알고 있는 경우 직접 입력할 수 있습니다. 키워드를 모르는 경우 부명령문 이름에 따라 Ctrl+스페이스바를 눌러 어디에서나 커서를 놓으면 부명령문에 사용 가능한 키워드 목록을 얻을 수 있습니다. 부명령문에 대한 자동 완성 제어를 표시합니다.

BARCHART 부명령문에서 PERCENT 키워드를 삭제하십시오.

Ctrl-스페이스바를 클릭합니다.

1. 빈도에 대해 **FREQ**라는 항목을 클릭합니다. 자동 완성 제어에서 항목을 클릭하면 현재 커서 위치에 삽입됩니다.

기본적으로 자동 완성 제어는 입력할 때 사용할 수 있는 항목을 묻는 메시지를 표시합니다. 예를 들어, 원형 도표를 막대도표와 함께 포함할 것입니다. 원형 도표는 별도의 부명령문으로 지정됩니다.

2. **FREQ** 키워드 다음에 Enter를 입력하고 포워드 슬래시(/)를 입력하여 부명령문 시작을 나타냅니다.

구문 편집기는 현재 명령의 부명령문 목록을 선택하라는 메시지를 표시합니다.

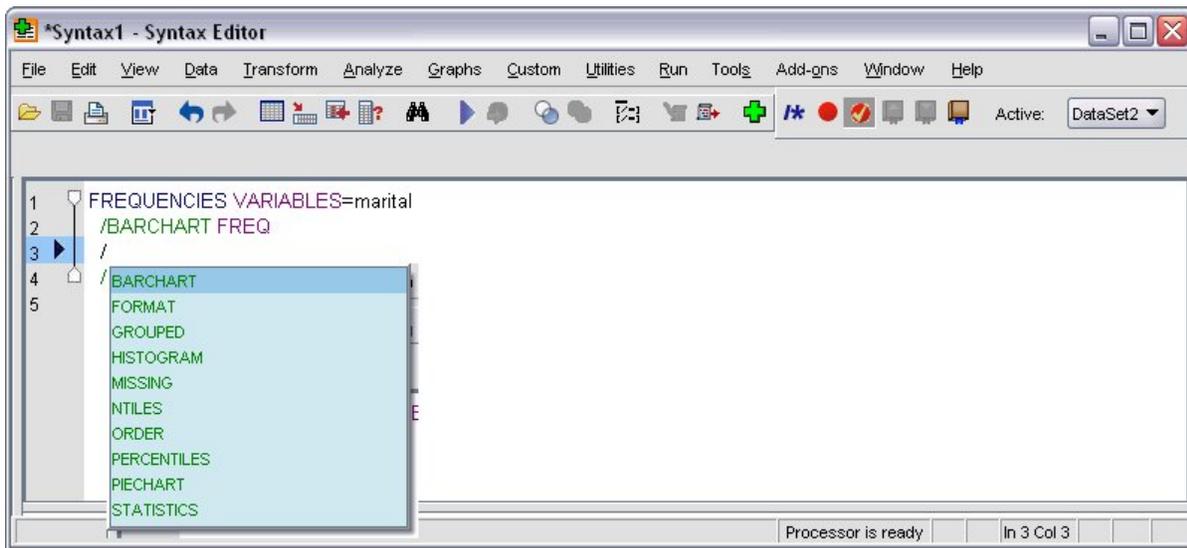


그림 71. 부명령문을 표시하는 자동 완성 제어

현재 명령에 대한 자세한 도움말을 보려면 F1 키를 클릭합니다. 이렇게 하면 현재 명령의 명령 구문 참조 정보로 바로 이동합니다.

구문 창에 표시되는 텍스트는 색상이 있다는 것을 알 수 있습니다. 색상 코딩을 사용하면 인식되는 항목만 색상이 표시되므로 인식되지 않는 항목을 신속하게 식별할 수 있습니다. 예를 들어, **FORMAT** 부명령문을 **FRMAT**로 잘못 적는다면 부명령문을 기본적으로 녹색으로 표시되지만 텍스트 **FRMAT**는 인식되지 않기 때문에 색상 없이 나타날 것입니다.

## 구문 파일 열기 및 실행

1. 저장된 구문 파일을 열려면 메뉴에서 다음을 선택합니다.

파일 > 열기 > 구문...

파일 열기 표준 대화 상자가 표시됩니다.

2. 구문 파일을 선택합니다. 구문 파일이 표시되지 않으면 **구문(\*.sps)**이 보기 원하는 파일 형식으로 선택되어 있는지 확인하십시오.
3. 열기를 클릭합니다.
4. 구문 창에서 실행 메뉴를 사용하여 명령을 실행합니다.

명령을 특정 데이터 파일에 적용하려면 명령을 실행하기 전에 데이터 파일이 열려 있거나, 열려 있지 않은 경우 데이터 파일을 여는 명령을 함께 실행해야 합니다. 데이터 파일을 연 대화 상자에서 이러한 유형의 명령을 붙여넣을 수 있습니다.

## 중단점 사용하기

중단점을 사용하면 구문 내의 특정 지점에서 명령 구문 실행을 중단할 수 있으며 준비가 되면 다시 실행할 수 있습니다. 따라서 명령 구문 작업의 임시 지점에서 FREQUENCIES 같은 데이터의 현재 상태에 대한 정보를 표시하는 결과나 데이터를 볼 수 있습니다. 중단점은 명령 내의 특정 줄이 아닌 명령 수준에서만 설정할 수 있습니다.

명령에 중단점을 삽입하려면

1. 명령과 관련된 텍스트의 왼쪽 영역 아무 곳이나 클릭합니다.

중단점은 마우스로 클릭한 위치에 관계 없이 영역의 명령 텍스트 왼쪽 및 명령 이름과 동일한 줄에서 빨간색 원으로 표시됩니다.

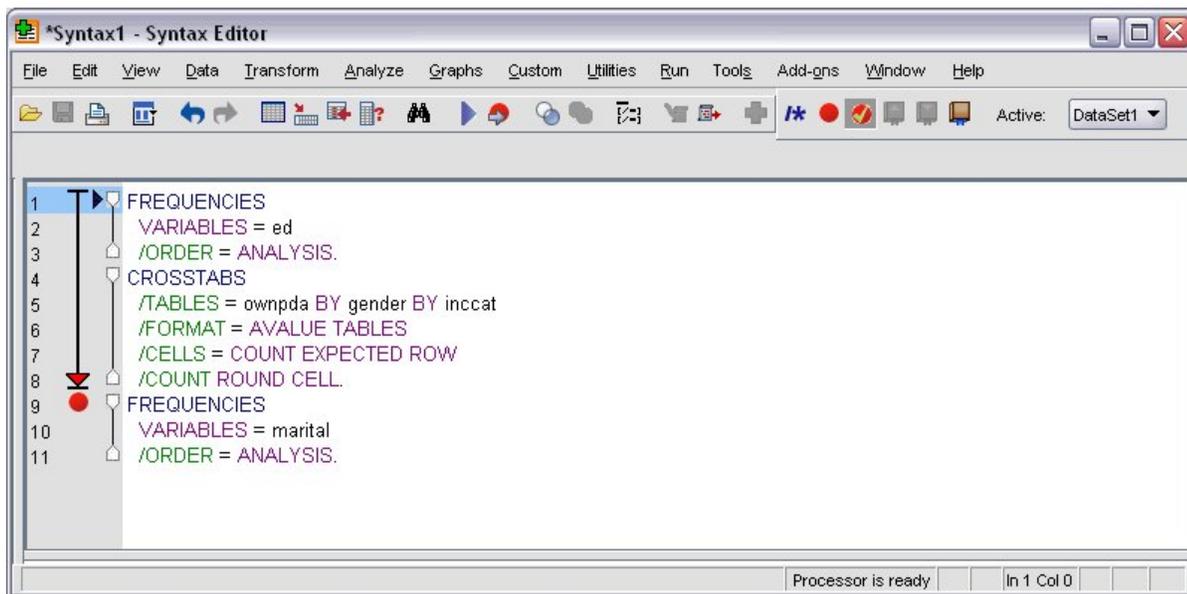


그림 72. 중단점에서 중지된 실행

중단점이 포함된 명령 구문을 실행하려면 중단점이 포함된 각 명령 앞에서 실행이 중지됩니다.

명령 텍스트 왼쪽의 아래를 향한 화살표는 구문 실행의 진행을 보여줍니다. 첫 번째 명령에서 마지막 명령 실행까지 영역에 걸쳐 있으며 특히 중단점이 포함되어 있는 명령 구문을 실행할 때 유용합니다.

중단점 이후의 실행을 다시 시작하려면

2. 구문 편집기의 메뉴에서 다음을 선택합니다.

**실행 > 계속**



---

## 제 8 장 데이터 값 수정

데이터가 항상 분석 또는 보고에 가장 적합한 형식으로 준비되어 있지는 않습니다. 예를 들어, 다음 작업을 수행해야 할 수 있습니다.

- 척도변수에서 범주형 변수를 작성합니다.
- 여러 개의 반응 범주를 단일 범주로 조합합니다.
- 기존 두 변수 간에 계산된 차이인 새로운 변수를 작성합니다.
- 두 날짜 사이의 시간 길이를 계산합니다.

이 장에서는 데이터 파일 *demo.sav*를 사용합니다. 자세한 정보는 91 페이지의 제 10 장 『표본 파일』 주제를 참조하십시오.

---

### 척도변수에서 범주형 변수 작성

*demo.sav* 데이터 파일에 있는 여러 개의 범주형 변수는 실제로 이 데이터 파일에 있는 척도변수에서 작성됩니다. 예를 들어, 소득범주 변수는 네 개의 가구소득 범주로 분류됩니다. 이 범주형 변수는 정수 값 1-4를 사용하여 \$25 미만, \$25-\$49, \$50-\$74, \$75 이상의 가구소득 범주(단위:1000)를 나타냅니다.

소득범주 범주형 변수 작성:

1. 데이터 편집기 창의 메뉴에서 다음을 선택합니다.

**변환 > 비주얼 구간화...**

초기 시각적 구간화 대화 상자에서 구간화된 새로운 변수를 작성할 척도변수 또는 순서변수를 선택합니다. 구간화란 둘 이상의 연속 값을 가져와서 같은 범주로 분류하는 것입니다.

시각적 구간화는 사용자가 적합한 구간화를 선택할 수 있도록 데이터 파일에 있는 실제 값을 사용하므로 먼저 데이터 파일을 읽어야 합니다. 데이터 파일에 케이스가 많을 경우 이 작업은 시간이 걸리므로 이 초기 대화 상자에서 읽을("스캔") 케이스 수를 제한할 수 있습니다. 표본 데이터 파일에는 이 작업이 필요하지 않습니다. 6,000개 이상의 케이스가 들어 있어도 이 개수의 케이스를 스캔하는 데 시간이 많이 걸리지 않습니다.

2. 변수 목록에서 구간화할 변수 목록으로 가구소득 [단위:1000] [가구소득]을 끌어 놓은 다음 **계속**을 클릭합니다.

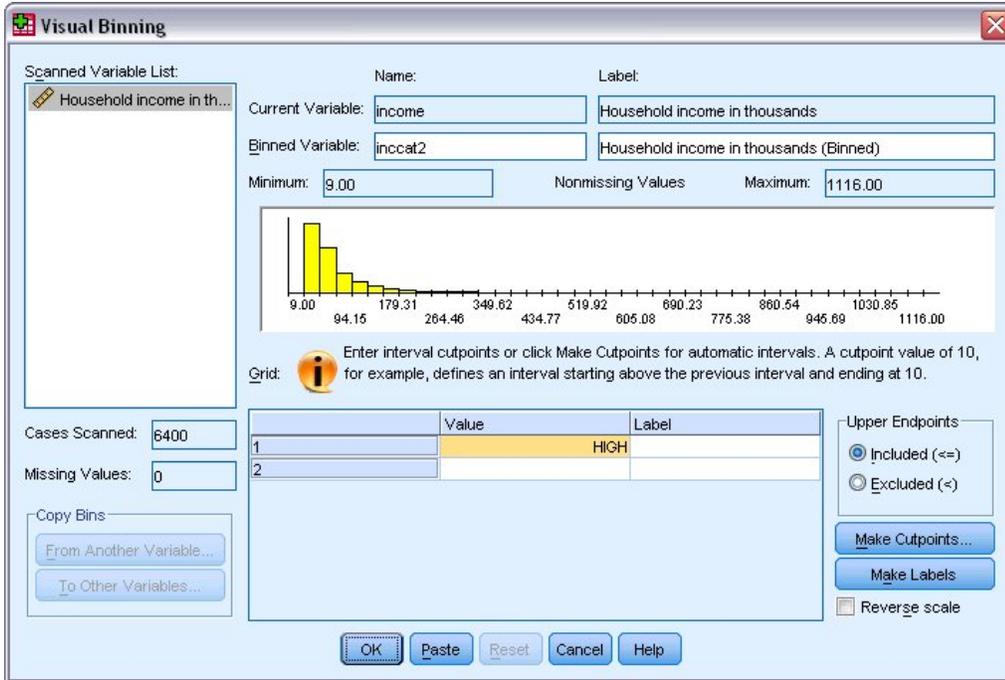


그림 73. 기본 시각적 구간화 대화 상자

3. 기본 시각적 구간화 대화 상자의 스캔된 변수 목록에서 가구소득 [단위:1000] [가구소득]을 선택합니다.

히스토그램은 선택한 변수의 분포(이 경우 비대칭)를 표시합니다.

4. 새로 구간화된 변수 이름으로 소득범주2를, 변수 레이블로 가구소득 범주 [단위: 1000]를 입력합니다.
5. 절단점 만들기를 클릭합니다.
6. 동일한 너비 구간을 선택합니다.
7. 처음 절단점 위치로 25, 절단점 수로 3, 절단점 너비로 25를 입력합니다.

구간화된 범주 수는 절단점 수에 1을 더한 값입니다. 그러므로 이 예제에서 새로 구간화된 변수는 모두 네 개의 범주를 갖게 됩니다. 처음 세 개 범주는 각각 25(천 단위) 범위를 포함하고 마지막 한 개 범주는 가장 큰 절단점 값 75(천 단위)보다 큰 모든 값을 포함합니다.

8. 적용을 클릭합니다.

현재 눈금에 표시된 값은 각 범주의 상한 엔드포인트로 정의된 절단점을 나타냅니다. 히스토그램의 수직선은 절단점의 위치를 나타냅니다.

기본적으로 이러한 절단점 값은 해당 범주에 포함되어 있습니다. 예를 들어, 첫 번째 값 25는 25보다 작거나 같은 모든 값을 포함합니다. 그러나, 이 예제에서는 25 미만, 25 - 49, 50 - 74 또는 75 이상에 해당하는 범주가 필요합니다.

9. 상한 엔드포인트에서 제외(<)를 선택합니다.

10. 그런 다음 레이블 만들기를 클릭합니다.

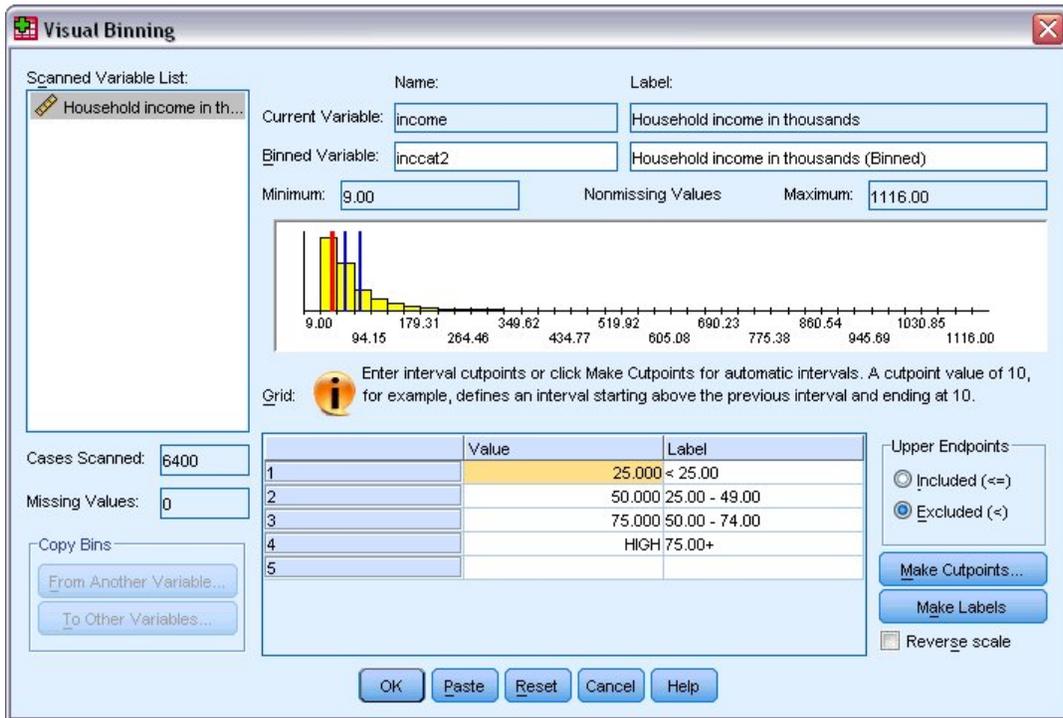


그림 74. 자동으로 생성된 값 레이블

각 범주에 대한 구체적인 값 레이블을 자동으로 생성합니다. 새로 구간화된 변수에 할당된 실제 값은 1로 시작하는 순차 정수이므로 이 값 레이블이 매우 유용할 수 있습니다.

또한, 눈금에 있는 절단점과 레이블을 수동으로 입력하거나 변경하고 히스토그램의 절단점 선을 끌어 놓아 절단점 위치를 변경할 뿐만 아니라 절단점 선을 히스토그램 밖으로 끌어서 절단점을 삭제할 수 있습니다.

11. 확인을 클릭하여 새로 구간화된 변수를 작성합니다.

새 변수가 데이터 편집기에 표시됩니다. 변수가 파일 끝에 추가되므로 데이터 보기의 오른쪽 끝에 있는 열과 변수 보기의 마지막 행에 표시됩니다.

## 새로운 변수 계산

광범위한 수학 함수를 사용하면 매우 복잡한 방정식에 따라 새로운 변수를 계산할 수 있습니다. 그러나, 이 예제에서는 기존 두 변수 값 차이인 새로운 변수를 간단히 계산하겠습니다.

*demo.sav* 데이터 파일에 반응자의 현재 연령에 대한 변수와 현재 회사에서의 근속연수에 대한 변수가 들어 있습니다. 그러나, 반응자가 해당 회사에 처음 입사했을 때의 연령에 대한 변수는 들어 있지 않습니다. 현재 연령과 현재 회사에서의 근속연수 간의 차이를 계산한 새로운 변수를 작성할 수 있으며, 이 변수는 반응자가 그 회사에 입사했을 때의 정확한 연령이 됩니다.

1. 데이터 편집기 창의 메뉴에서 다음을 선택합니다.

**변환 > 변수 계산...**

2. 대상 변수에 jobstart를 입력합니다.
3. 소스 변수 목록에서 연령[단위:세] [연령]을 선택하고 화살표 단추를 클릭하여 숫자표현식 입력란에 복사합니다.
4. 대화 상자의 계산기에 있는 빼기(-) 단추를 클릭하거나 키보드의 빼기 키를 클릭합니다.
5. 현직장 근무년수 [근무년수]를 선택하고 화살표 단추를 클릭하여 표현식에 복사합니다.

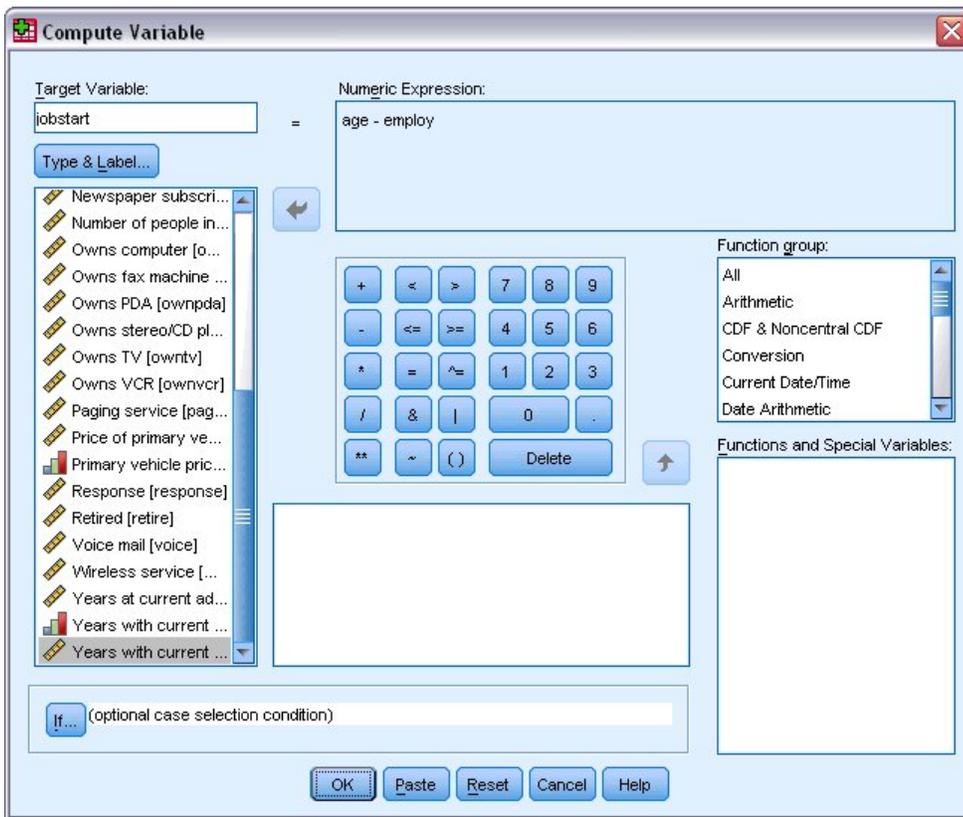


그림 75. 변수 계산 대화 상자

참고: 올바른 직원 변수를 선택하도록 주의하십시오. 또한 코딩변경된 범주형 변수 버전이 있습니다. 이것은 귀하가 원하는 변수가 아닙니다. 숫자표현식은 연령-*empcat*이 아닌 연령-근무년수여야 합니다.

6. 확인을 클릭하여 새로운 변수를 계산합니다.

새 변수가 데이터 편집기에 표시됩니다. 변수가 파일 끝에 추가되므로 데이터 보기의 오른쪽 끝에 있는 열과 변수 보기의 마지막 행에 표시됩니다.

## 표현식에 함수 사용

표현식에 미리 정의된 함수를 사용할 수도 있습니다. 70개 이상의 기본 함수를 사용할 수 있으며, 함수의 예는 다음과 같습니다.

- 산술 함수
- 통계 함수
- 분포 함수
- 논리 함수
- 날짜와 시간 통합 함수 및 추출 함수
- 결측값 함수
- 교차 케이스 함수
- 문자열 함수

함수는 산술 연산 그룹, 통계 메트릭 계산 그룹 등 논리적으로 구별되는 그룹으로 구성되어 있습니다. 편의를 위해  $TIME$ (현재 날짜 및 시간)과 같은 일반적으로 사용되는 몇 가지 시스템 변수도 적합한 함수 그룹에 포함됩니다.

### 표현식에 함수 붙여넣기

함수를 표현식에 붙여넣기:

1. 표현식에서 함수가 표시될 위치에 커서를 놓습니다.
2. 함수 그룹 목록에서 적절한 그룹을 선택합니다. 모두 레이블이 표시된 그룹은 사용 가능한 모든 함수 및 시스템 변수 목록을 제공합니다.
3. 함수 목록 및 특수변수 목록의 함수를 두 번 클릭하거나, 함수를 선택하고 함수 그룹 목록 옆 화살표를 클릭합니다.

함수가 표현식에 삽입됩니다. 표현식의 일부를 선택한 다음 함수를 삽입하면 선택한 표현식 부분이 함수의 첫 번째 인수로 사용됩니다.

### 표현식에서 함수 편집

인수를 입력할 때까지 함수가 완료되지 않고 붙여넣은 함수에 물음표가 표시됩니다. 물음표 개수는 해당 함수를 완료하는 데 필요한 최소 인수 개수를 나타냅니다.

1. 붙여넣은 함수에서 물음표를 선택합니다.
2. 인수를 입력합니다. 인수가 변수 이름이면 변수 목록에서 붙여넣을 수 있습니다.

## 조건식 사용

조건식(논리 표현식이라고도 함)을 사용하여 선택한 케이스 서브세트에 변환을 적용할 수 있습니다. 조건식은 각 케이스에 대해 True, False의 값 또는 결측값을 반환합니다. 조건식 결과가 True이면 해당 케이스에 변환이 적용됩니다. 조건식 결과가 False 또는 결측값이면 케이스에 변환이 적용되지 않습니다.

조건식 지정하기:

1. 변수 계산 대화 상자에서 **조건**을 클릭합니다. 케이스 조건 대화 상자가 열립니다.
2. 다음 조건을 만족하는 케이스 포함을 선택합니다.
3. 조건식을 입력합니다.

대부분의 조건식에는 아래와 같이 하나 이상의 관계 연산자가 있습니다.

연령>=21

또는

가구소득\*3<100

첫 번째 예제에서 연령 [연령]의 값이 21인 케이스만 선택됩니다. 두 번째 예제에서는 가구소득 [단위:1000] [가구소득]에 3을 곱한 값이 100보다 작은 케이스가 선택되어야 합니다.

아래와 같이 논리 연산자를 사용하여 둘 이상의 조건식을 연결할 수도 있습니다.

연령>=21 | 교육수준>=4

또는

가구소득\*3<100 & 교육수준=5

첫 번째 예제에서 연령 [연령] 조건 또는 피교육수준 [교육수준] 조건을 만족하는 케이스가 선택됩니다. 두 번째 예제에서는 가구소득 [단위:1000] [가구소득] 조건과 피교육수준 [교육수준] 조건을 모두 만족하는 케이스가 선택되어야 합니다.

---

## 날짜 및 시간을 사용한 작업

일반적으로 날짜 및 시간을 사용하는 여러 작업은 날짜 및 시간 마법사를 사용하여 쉽게 수행할 수 있습니다. 이 마법사를 사용하여 수행할 수 있는 작업은 다음과 같습니다.

- 날짜 또는 시간이 포함된 문자열 변수에서 날짜/시간변수를 작성합니다.
- 날짜 또는 시간의 여러 부분이 포함된 변수를 병합하여 날짜/시간변수를 구성합니다.
- 두 날짜/시간변수를 더하거나 빼는 등 날짜/시간변수에서 값을 더하거나 뺍니다.
- 날짜 또는 시간변수의 일부, 예를 들어 mm/dd/yyyy 형식의 날짜/시간변수에서 일자를 추출합니다.

이 섹션의 예제에서는 데이터 파일 *upgrade.sav*를 사용합니다. 자세한 정보는 91 페이지의 제 10 장 『표본 파일』 주제를 참조하십시오.

날짜 및 시간 마법사 사용하기:

1. 메뉴에서 다음을 선택합니다.

## 변환 > 날짜 및 시간 마법사...



그림 76. 날짜 및 시간 마법사 안내 화면

날짜 및 시간 마법사 안내 화면에는 일반 작업 세트가 표시되어 있습니다. 현재 데이터에 적용되지 않는 작업은 사용할 수 없습니다. 예를 들어, *upgrade.sav* 데이터 파일에는 문자열 변수가 없으므로 문자열에서 날짜 변수를 작성하는 작업은 사용할 수 없습니다.

IBM SPSS Statistics에서 날짜 및 시간을 처음 사용하는 경우 **날짜 및 시간을 표시하는 방법**을 선택하고 다음을 클릭합니다. 그러면 날짜/시간변수에 대한 간단한 개요와 도움말 단추를 통해 자세한 내용을 볼 수 있는 링크를 제공하는 화면이 나타납니다.

### 두 날짜 사이의 시간 길이 계산

날짜와 관련된 가장 일반적인 작업 중 하나는 두 날짜 사이의 시간 길이를 계산하는 것입니다. 예를 들어, 각 고객이 마지막으로 업그레이드를 구매한 이후의 연도 수를 파악하여 업그레이드 라이선스 구매를 분석하는 데 관심이 있는 소프트웨어 회사가 있다고 가정합니다. *upgrade.sav* 데이터 파일에는 각 고객이 마지막으로 업그레이드를 구매한 날짜에 대한 변수는 있지만 그 이후에 경과한 연도 수에 대한 변수는 없습니다. 이 경우 마지막 업그레이드 날짜와 다음 제품 릴리스 날짜 사이의 시간 길이(연도 수)를 새로운 변수로 사용하여 이 수량을 측정할 수 있습니다.

두 날짜 사이의 시간 길이 계산하기:

1. 날짜 및 시간 마법사 안내 화면에서 **날짜 및 시간으로 계산**을 선택하고 다음을 클릭합니다.
2. **두 날짜 사이의 시간 단위 수 계산**을 선택하고 다음을 클릭합니다.

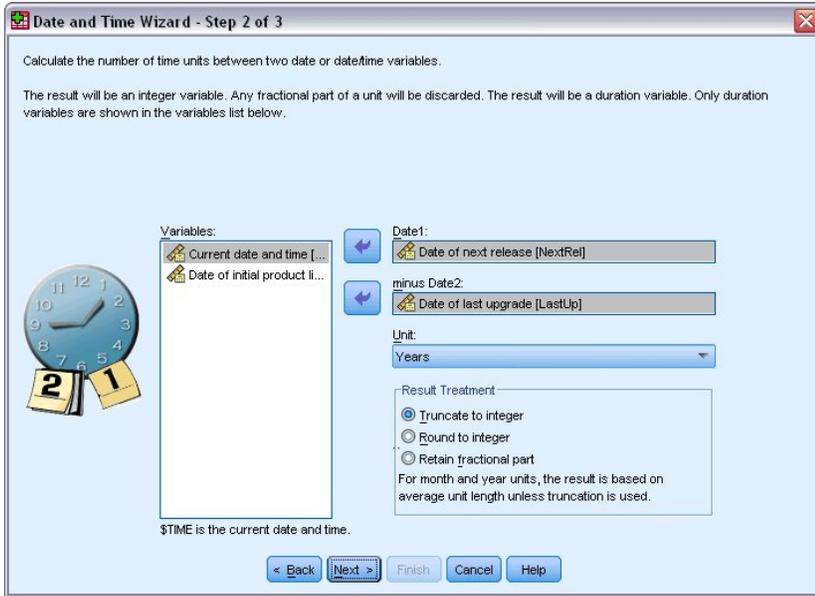


그림 77. 두 날짜 사이의 시간 길이 계산: 2단계

3. 2단계에서 날짜1에 다음 버전이 출시될 날짜를 선택합니다.
4. 날짜2에 마지막 업그레이드 날짜를 선택합니다.
5. 단위로 **년**을 선택하고 결과 처리를 위해 **정수로 내림**을 선택합니다. (이것은 기본값 선택입니다.)
6. **다음**을 클릭합니다.
7. 3단계에서 결과변수 이름에 *YearsLastUp*를 입력합니다. 결과변수 이름은 기존 변수 이름과 같을 수 없습니다.
8. 결과변수에 대한 레이블로 마지막 업그레이드 이후의 연도 수를 입력합니다. 결과변수에 대한 변수 레이블은 선택 사항입니다.
9. 기본값 선택인 **지금 변수 생성**을 그대로 두고 **마침**을 클릭하여 새로운 변수를 작성합니다.

데이터 편집기에 표시된 새로운 변수 업그레이드날짜는 두 날짜 사이의 연도 수(정수)입니다. 연도 수의 분수 부분은 잘립니다.

## 날짜에 기간 더하기

날짜에 10일 또는 12개월 등의 기간을 더하거나 뺄 수 있습니다. 이전 섹션에서 사용한 소프트웨어 회사를 계속 예로 들어 각 고객의 최초 기술 지원 계약 종료 날짜를 결정해 봅시다. *upgrade.sav* 데이터 파일에는 지원 계약 연도 수에 대한 변수 및 최초 구매 날짜에 대한 변수가 포함되어 있습니다. 따라서 구매 날짜에 지원 연도 수를 더하여 최초 지원 종료 날짜를 결정할 수 있습니다.

날짜에 기간을 더하기:

1. 날짜 및 시간 마법사 안내 화면에서 **날짜 및 시간으로 계산**을 선택하고 **다음**을 클릭합니다.
2. 1단계에서 **날짜에서 기간을 더하거나 빼기**를 선택하고 **다음**을 클릭합니다.

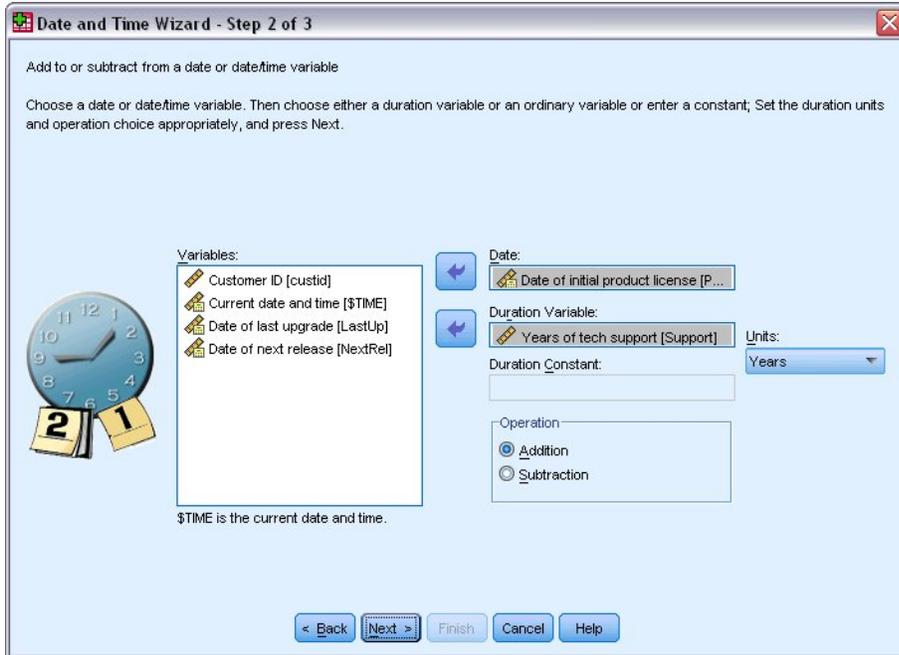


그림 78. 날짜에 기간 더하기: 2단계

3. 날짜에 초기 생산품 라이선스의 날짜를 선택합니다.
4. 2단계에서 기간 변수에 기술지원 가능 년수를 선택합니다.

기술지원 가능 년수는 숫자변수이므로 이 변수를 기간으로 더할 경우에는 사용할 단위를 표시해야 합니다.

5. 단위 드롭다운 목록에서 년을 선택합니다.
6. 다음을 클릭합니다.
7. 3단계에서 결과변수 이름에 *SupEndDate*를 입력합니다. 결과변수 이름은 기존 변수 이름과 같을 수 없습니다.
8. 결과변수의 레이블로 지원 종료 날짜를 입력합니다. 결과변수에 대한 변수 레이블은 선택 사항입니다.
9. 마침을 클릭하여 새로운 변수를 작성합니다.

새 변수가 데이터 편집기에 표시됩니다.



---

## 제 9 장 데이터 정렬 및 선택

데이터 파일은 항상 특정 요구에 맞게 이상적인 형태로 구성되지는 않습니다. 분석을 목적으로 데이터를 준비하기 위하여 다음과 같은 기능들을 포함한 광범위한 파일 변환물 중에서 선택할 수 있습니다:

- **데이터 정렬.** 하나 이상의 변수 값을 기준으로 케이스를 정렬할 수 있습니다.
- **케이스 서브세트 선택.** 케이스 서브세트로 분석을 제한하거나 여러 서브세트에서 동시에 분석을 수행할 수 있습니다.

이 장의 예제에서는 *demo.sav* 데이터 파일을 사용합니다. 자세한 정보는 91 페이지의 제 10 장 『표본 파일』 주제를 참조하십시오.

---

### 데이터 정렬

케이스 정렬은 (데이터 파일의 행 정렬) 종종 유용하며 특정 유형의 분석에 때때로 필요합니다.

여러 정렬변수 값에 따라 데이터 파일의 케이스 시퀀스 바꾸기:

1. 메뉴에서 다음을 선택합니다.

    데이터 > 케이스 정렬...

    케이스 정렬 대화 상자가 표시됩니다.

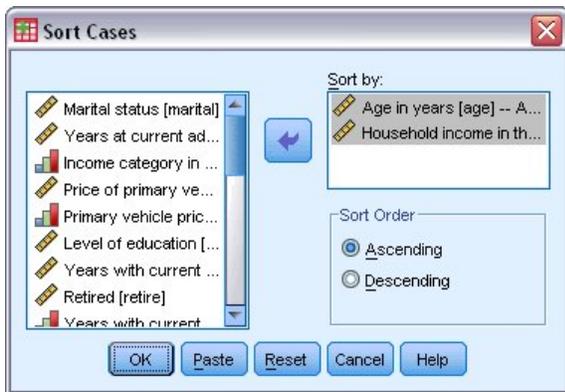


그림 79. 케이스 정렬 대화 상자

2. 연령 [연령] 및 가구소득 범주[단위:1000][소득] 변수를 목록별 정렬에 추가합니다.

다중 정렬변수를 선택할 때는 목록별 정렬에 나타나는 순서가 케이스 정렬 순서를 결정합니다. 이 예에서 목록별 정렬의 항목을 기반으로 케이스는 연령 [연령] 범주 내에서 가구소득 범주[단위:1000][소득] 값에 의해 정렬됩니다. 문자열 변수의 경우 정렬 순서에서 대문자는 소문자보다 먼저 나옵니다. (예를 들어 문자열 값 Yes는 정렬 순서에서 yes보다 먼저 나옵니다.)

## 분할 파일 처리

분석을 위해 데이터 파일을 별개의 그룹으로 분할:

1. 메뉴에서 다음을 선택합니다.

데이터 > 분할 파일...

분할 파일 대화 상자가 표시됩니다.

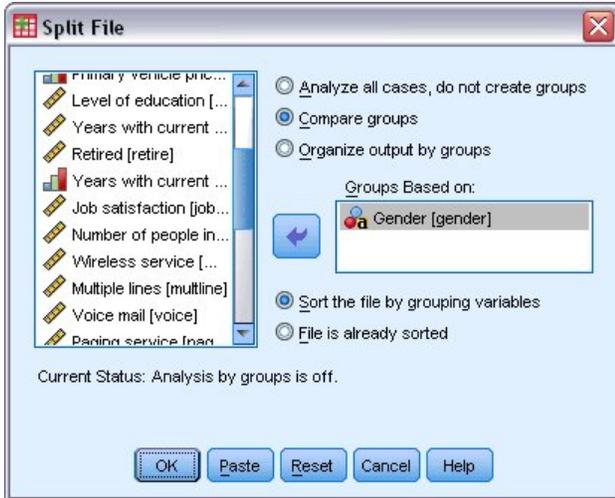


그림 80. 분할 파일 대화 상자

2. **그룹 비교** 또는 **각 그룹별로 결과를 나타냄**을 선택합니다. (이 단계 뒤에 나오는 예제들은 이 두 옵션 사이의 차이점을 보여 줍니다.)
3. **성별 [성별]**을 선택하여 이들 변수에 대해 파일을 별개의 그룹으로 분할합니다.

숫자 변수, 단문자열 변수, 장문자열 변수를 그룹화 변수로 사용할 수 있습니다. 그룹화 변수에 의해 정의되는 각 하위 그룹에 대해 개별 분석이 수행됩니다. 다중 그룹화 변수를 선택할 때는 분할 그룹화 변수 목록에 나타나는 순서가 케이스를 그룹화하는 방법을 결정합니다.

그룹 비교를 선택할 경우 모든 분할 파일 그룹의 결과가 빈도분석 프로시저에 의해 생성된 다음의 요약 통계 표에 나타난 것처럼 동일한 표에 포함될 것입니다.

**Statistics**

Household income in thousands

Female	N	Valid	3179
		Missing	0
	Mean		68.7798
	Median		44.0000
	Std. Deviation		75.73510
Male	N	Valid	3221
		Missing	0
	Mean		70.1608
	Median		45.0000
	Std. Deviation		81.56216

그림 81. 단일 피벗 표가 있는 분할 파일 결과

각 그룹별로 결과를 나타냄을 선택하고 빈도분석 프로시저를 실행할 경우 두 개의 피벗 표가 생성됩니다(한 개의 여자 피벗 표 및 한 개의 남자 피벗 표)

**Statistics<sup>a</sup>**

Household income in thousands

N	Valid	3179
	Missing	0
Mean		68.7798
Median		44.0000
Std. Deviation		75.73510

a. Gender = Female

그림 82. 여자 피벗 표가 있는 분할 파일 결과

**Statistics<sup>a</sup>**

Household income in thousands

N	Valid	3221
	Missing	0
Mean		70.1608
Median		45.0000
Std. Deviation		81.56216

a. Gender = Male

그림 83. 남자 피벗 표가 있는 분할 파일 결과

## 분할 파일 처리의 케이스 정렬

분할 파일 프로시저는 그룹화 변수중 하나에 대해 다른 값을 만나게 될 때마다 새 하위 그룹을 생성합니다. 따라서, 분할 파일 처리를 호출하기 전 그룹화 변수 값을 기반으로 케이스를 정렬하는 것이 중요합니다.

기본적으로 분할 파일은 그룹화 변수 값을 기반으로 데이터 파일을 자동으로 정렬합니다. 파일이 적절한 순서대로 이미 정렬되어 있는 경우 **파일이 이미 정렬되어 있음**을 선택하면 처리 시간이 단축될 수 있습니다.

## 분할 파일 처리 켜기 및 끄기

분할 파일 처리를 호출한 후 이것을 끄지 않으면 세션 나머지 부분에 대해서도 계속 효력을 갖습니다.

- 모든 케이스 분석. 이 옵션은 분할 파일 처리를 끕니다.
- 그룹 비교 및 각 그룹별로 결과를 나타냄. 이 옵션은 분할 파일 처리를 켭니다.

분할 파일 처리가 효력을 나타낼 경우 **분할 파일 켜기**의 메시지가 애플리케이션 창 하단의 상태 표시 줄에 표시됩니다.

---

## 케이스 서브세트 선택

변수와 복잡한 표현식을 포함한 범주를 기준으로 분석을 특정 하위 그룹으로 제한할 수 있습니다. 케이스의 무작위 표본을 선택할 수도 있습니다. 부그룹을 정의하는데 사용되는 기준은 다음과 같습니다.

- 변수 값과 범위
- 날짜와 시간 범위
- 케이스 (행) 번호
- 산술 표현식
- 논리 표현식
- 함수

분석을 위해 케이스의 서브세트 선택:

1. 메뉴에서 다음을 선택합니다.

**데이터 > 케이스 선택...**

케이스 선택 대화 상자가 열립니다.

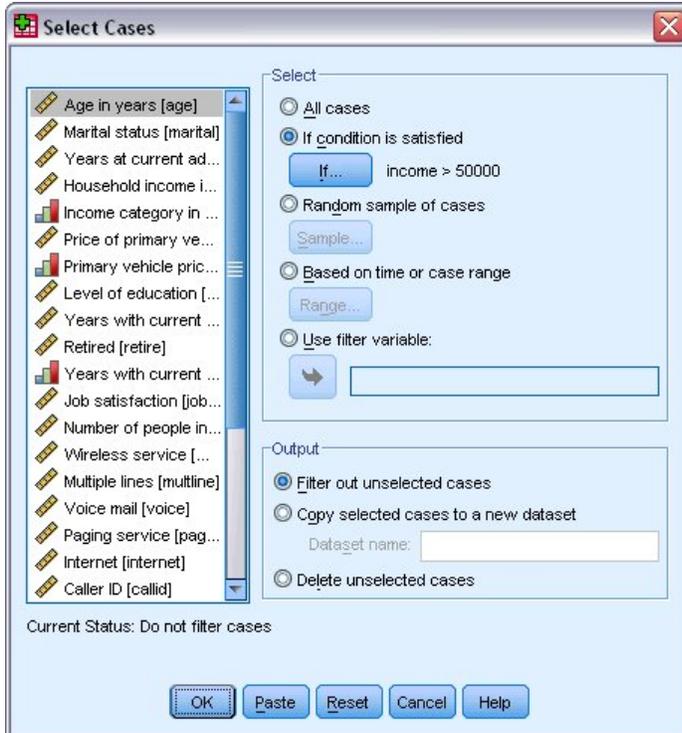


그림 84. 케이스 선택 대화 상자

## 조건식을 기준으로 케이스 선택

조건식을 기반으로 케이스 선택:

1. 조건을 만족하는 케이스를 선택하고 케이스 선택 대화 상자에서 조건을 클릭합니다.

케이스 조건 선택 대화 상자가 열립니다.

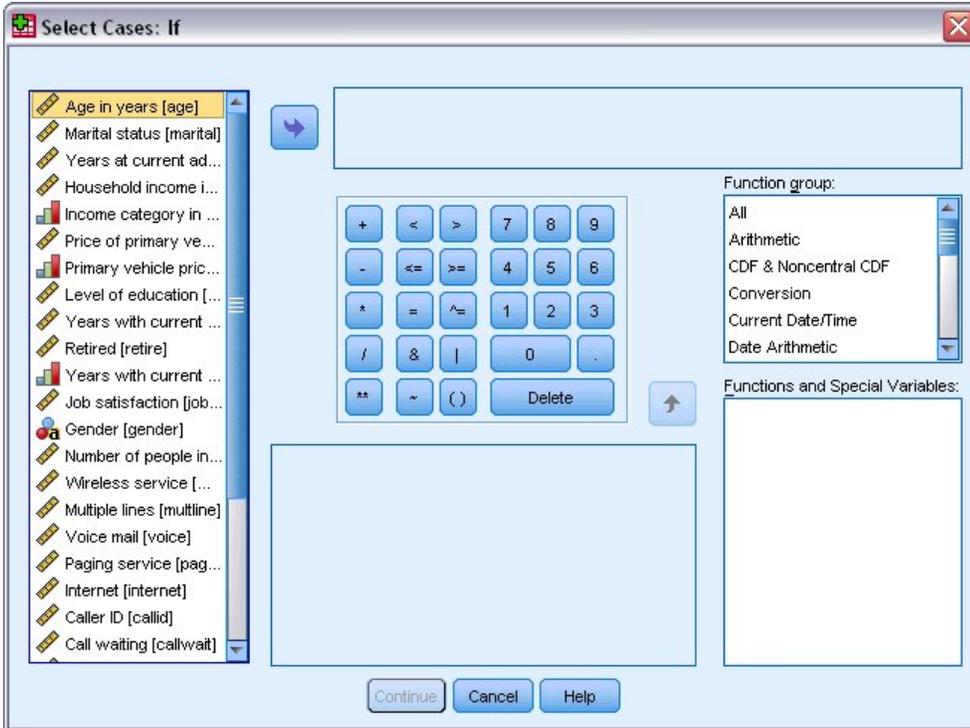


그림 85. 케이스 조건 선택 대화 상자

조건식은 기존의 변수 이름, 상수, 산술 연산자, 논리 연산자, 관계 연산자 및 함수를 사용할 수 있습니다. 결과 창의 텍스트와 마찬가지로 입력란 안에 표현식을 입력하고 편집할 수 있습니다. 또 계산기 패드, 변수 목록, 함수 목록을 사용하여 표현식 안에 요소를 붙여넣을 수 있습니다. 자세한 정보는 77 페이지의 『조건식 사용』 주제를 참조하십시오.

## 케이스의 무작위 표본 선택

무작위 표본 구하기:

1. 케이스 선택 대화 상자에서 **케이스의 무작위 표본**을 선택합니다.
2. **표본**을 클릭합니다.

케이스 무작위 표본 선택 대화 상자가 열립니다.

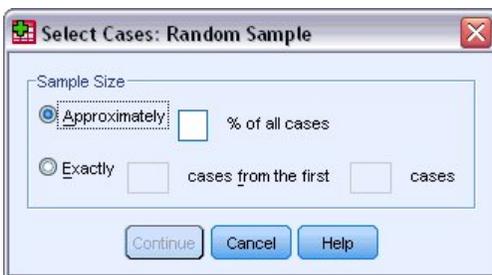


그림 86. 케이스 무작위 표본 선택 대화 상자

표본 결과에 대해 다음 선택 중 하나를 선택할 수 있습니다:

- **대략.** 사용자가 지정한 퍼센트. 이 옵션은 대략적으로 지정된 케이스 퍼센트를 갖는 무작위 표본을 작성합니다.
- **정확하게.** 사용자가 지정한 케이스 수입니다. 표본을 작성할 케이스 수를 지정해야 합니다. 이 케이스 수는 데이터 파일의 총 케이스 수보다 작거나 같아야 합니다. 이 수가 데이터 파일의 총 수를 넘을 경우에는 상대적으로 요청한 수보다 적은 케이스가 표본에 포함됩니다.

## 시간 범위 또는 케이스 범위 선택

날짜, 시간 또는 관측값(행) 번호를 기반으로 케이스 범위 선택:

1. 시간 또는 케이스 범위를 기준으로 선택하고 케이스 선택 대화 상자에서 범위를 클릭합니다.

이렇게 하면 케이스 범위 선택 대화 상자가 열리며 여기에서 광범위한 관측값(행) 번호를 선택할 수 있습니다.

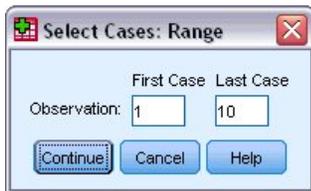


그림 87. 케이스 범위 선택 대화 상자

- **첫번째 케이스.** 범위에 대한 시작 날짜 및/또는 시간 값을 입력합니다. 정의된 날짜 변수가 없을 경우 시작 관측값 번호를 입력합니다(분할 파일이 켜져 있지 않은 경우 데이터 편집기의 행 번호). 마지막 케이스 값을 지정하지 않을 경우 시작 날짜/시작 시간에서 종료 시간까지 모든 케이스 계열이 선택됩니다.
- **마지막 케이스.** 범위에 대한 종료 날짜 및/또는 시간 값을 입력합니다. 정의된 날짜 변수가 없을 경우 종료 관측값 번호를 입력합니다(분할 파일이 켜져 있지 않은 경우 데이터 편집기의 행 번호). 첫 번째 케이스 값을 지정하지 않을 경우 시작 시간에서 종료 날짜/종료 시간까지 모든 케이스 계열이 선택됩니다.

날짜 변수가 정의된 시계열 데이터의 경우 정의된 날짜변수를 기준으로 날짜 범위 및/또는 시간 범위를 선택할 수 있습니다. 각 케이스는 서로 다른 시간의 관측값을 나타내며 파일은 연대기적 순서로 정렬됩니다.



그림 88. 케이스 범위 선택 대화 상자(시계열)

시계열 데이터의 날짜 변수 작성:

2. 메뉴에서 다음을 선택합니다.

데이터 > 날짜 정의...

### 선택하지 않은 케이스 처리

선택되지 않은 케이스를 처리하기 위한 방법으로 다음 옵션 중 하나를 선택할 수 있습니다.

- **선택하지 않은 케이스 필터.** 선택되지 않은 케이스가 분석에 포함되지 않지만 데이터 세트에는 계속 남아 있습니다. 세션을 진행하는 동안 나중에라도 필터링을 해제하면 선택되지 않은 케이스를 사용할 수 있습니다. 무작위 표본을 선택하거나 조건식을 기반으로 케이스를 선택하는 경우 이 옵션을 사용하면 선택된 케이스의 값은 1이고 선택되지 않은 케이스의 값은 0인 *filter\_\$*라는 변수가 생성됩니다.
- **새 데이터 세트에 선택한 케이스 복사.** 선택된 케이스가 새로운 데이터 세트에 복사되고 원래 데이터 세트는 그대로 유지됩니다. 선택되지 않은 케이스는 새로운 데이터 세트에 포함되지 않으며 원래 데이터 세트에 원래 상태 그대로 남습니다.
- **선택하지 않은 케이스 삭제.** 선택되지 않은 케이스가 데이터 세트에서 삭제됩니다. 삭제된 케이스를 복구하려면 파일의 변경 내용을 저장하지 않은 채 파일을 닫고 다시 열어야 합니다. 데이터 파일의 변경 내용을 저장하면 삭제된 케이스를 복구할 수 없습니다.

참고: 선택되지 않은 케이스를 삭제하고 파일을 저장하면 케이스를 복구할 수 없습니다.

---

## 케이스 선택 상태

케이스 서브세트를 선택했으나 선택하지 않은 케이스를 버리지 않을 경우에는 선택하지 않은 케이스가 행 번호에 대각선 또는 슬래시(/)가 그어진 상태로 데이터 편집기에 표시됩니다.

1
2
3
4
5
6
7
8
9
10
11

그림 89. 케이스 선택 상태

## 제 10 장 표본 파일

제품에 설치된 표본 파일은 설치 디렉토리 표본 추출 하위 디렉토리에서 찾으실 수 있습니다. 영어, 프랑스어, 독일어, 이탈리아어, 일본어, 한국어, 폴란드어, 러시아어, 중국어 간체, 스페인어 및 중국어 번체 등의 각 언어에 대해 Samples 하위 디렉터리 내에 별도의 폴더가 있습니다.

모든 언어에서 모든 표본 파일을 사용할 수 있는 것은 아닙니다. 특정 언어로 표본 파일을 사용할 수 없는 경우 해당 언어 폴더에 영어 버전의 표본 파일이 들어 있습니다.

### 설명

이것은 문서 전반에서 다양한 예제에 사용된 표본 파일에 대한 요약 설명입니다.

- **accidents.sav.** 이것은 주어진 지역에서 자동차 사고에 대한 연령 및 성별 위험 요인을 연구하는 보험회사와 관련이 있는 가설 데이터 파일입니다. 각 케이스는 연령 범주 및 성별의 교차 분류에 대응합니다.
- **adl.sav.** 이것은 뇌졸중 환자에 대해 제시된 치료 유형의 장점을 판단하려는 노력과 관련이 있는 가설 데이터 파일입니다. 의사는 여성 뇌졸중 환자를 두 그룹 중 하나에 무작위 할당합니다. 첫 번째로 받은 표준 물리치료, 두 번째로 받은 추가 정서적 치료. 치료 후 석달 동안 각 환자가 일상생활에서 흔한 활동들을 수행하는 기능을 순서 변수로써 스코어를 매겼습니다.
- **advert.sav.** 이것은 광고에 소비된 돈과 그 결과로써 판매에 소비된 돈 사이의 관계를 조사하려는 유통업체의 시도와 관련이 있는 가설 데이터 파일입니다. 이를 위해 과거 매출 수치와 관련 광고비를 수집했으며,
- **aflatoxin.sav.** 이것은 옥수수 작물에 대해서 농작물마다 농도가 크게 다른 독성 물질인 아플라톡신 검정과 관련이 있는 가설 데이터 파일입니다. 어느 곡물 가공업자가 8가지 농작물 중에서 16개 표본을 얻어 10억분의 1단위로 아플라톡신 레벨을 측정하였습니다.

- **anorectic.sav** 연구원들은 식욕 감퇴/폭식증 등 표준적인 증상에 대한 연구와 병행하여<sup>1</sup> 알려진 섭식 장애를 갖고 있는 55명의 젊은이에 대한 연구를 수행하였습니다. 각 환자를 4년 동안 네 차례 진찰하여 총 220개의 관측값을 얻었습니다. 환자들을 대상으로 관찰하는 동안 16개 증상 각각에 대해 스코어를 매겼습니다. 환자 71번의 2회차, 환자 76번의 2회차, 환자 47번의 3회차 진료에 대한 증상 스코어가 누락되었으며 나머지 217개 관측값은 유효합니다.
- **bankloan.sav**. 이것은 은행이 기본 대부율을 낮추려는 시도와 관련이 있는 가설 데이터 파일입니다. 파일에는 과거 및 장래 전망 있는 850명의 고객에 대한 금융 정보 및 인구 통계학 정보가 들어 있습니다. 처음 700개의 케이스는 이전에 대출을 받은 고객입니다. 마지막 150 케이스는 은행이 신용 위험이 낮거나 높은 것으로 분류할 필요가 있는 장래 전망 있는 고객들입니다.
- **bankloan\_binning.sav**. 이것은 과거 5,000명의 고객들에 관한 금융 정보 및 인구 통계학 정보가 들어있는 가설 데이터 파일입니다.
- **behavior.sav**. 고전적인 예제<sup>2</sup>에서 52명의 학생들에게 부탁하여 15가지 상황과 15가지 행동의 조합에 대해 0 = "매우 적절함"에서 9 = "매우 부적절함"까지 10단계 척도로 순위를 매기도록 요청했습니다. 각 응답자에 대해 평균을 구한 값은 상이도로 간주됩니다.
- **behavior\_ini.sav**. 이 데이터 파일에는 *behavior.sav*에 대한 2차원적 솔루션의 초기 설정이 들어 있습니다.
- **brakes.sav**. 이것은 성능이 높은 자동차용 디스크 브레이크 생산 공장의 품질 관리와 관련이 있는 가설 데이터 파일입니다. 데이터 파일에는 생산 기계 8대 각각의 16개 디스크에 대한 지름 치수가 들어 있습니다. 브레이크의 대상 지름은 322mm입니다.
- **breakfast.sav**. 고전적인 연구<sup>3</sup>에서 21명의 Wharton School MBA 학생들과 그 배우자들에게 15가지 아침 식사 메뉴에 선호도에 따라 1부터 15까지 순위를 매기도록 부탁했습니다. 1은 '가장 좋아함'을 의미하고 15는 '가장 싫어함'을 의미합니다. "전반적인 선호"에서 "음료수만 포함된 가벼운 식사"에 이르는 서로 다른 여섯 가지 시나리오에 따라 응답자들의 선호도를 기록했습니다.
- **breakfast-overall.sav**. 이 데이터 파일에는 최초의 시나리오 "전반적인 선호도"에 대한 아침 식사 메뉴 중 좋아하는 음식이 들어 있습니다.
- **broadband\_1.sav**. 이것은 지역별 국가 광대역 서비스 가입자 수가 들어 있는 가설 데이터 파일입니다. 데이터 파일에는 4년 동안 85개 지역에 대한 매달 가입자 수가 들어 있습니다.
- **broadband\_2.sav**. 이 데이터 파일은 *broadband\_1.sav*와 동일하나 추가 3달에 대한 데이터가 들어 있습니다.

---

1. Van der Ham, T., J. J. Meulman, D. C. Van Strien, and H. Van Engeland. 1997. Empirically based subgrouping of eating disorders in adolescents: A longitudinal perspective. *British Journal of Psychiatry*, 170, 363-368.

2. Price, R. H., and D. L. Bouffard. 1974. Behavioral appropriateness and situational constraints as dimensions of social behavior. *Journal of Personality and Social Psychology*, 30, 579-586.

3. Green, P. E., and V. Rao. 1972. *Applied multidimensional scaling*. Hinsdale, Ill.: Dryden Press.

- **car\_insurance\_claims.sav.** 자동차 손해배상 청구와 관련하여 <sup>4</sup> 외에서 제시 및 분석된 데이터 세트입니다. 평균 청구 금액은 평균 종속 변수를 보험 계약자의 연령, 차종 및 차량 연식과 관련시켜 설명하는 역 연결 함수를 사용하여 감마 분포로써 모델링할 수 있습니다. 기록에 남긴 청구 개수는 척도 가중치로 사용될 수 있습니다.
- **car\_sales.sav.** 이 데이터 파일은 다양한 메이커와 차량 모형에 대한 예상 판매량, 정가 및 물리적 지정 사항을 포함합니다. 정가와 물리적 지정 사항은 *edmunds.com*과 제조업체 사이트에서 번갈아 입수하였습니다.
- **car\_sales\_upprepared.sav.** *car\_sales.sav*의 수정된 버전으로서 변환된 버전의 필드가 포함되지 않습니다.
- **carpet.sav.** 자주 인용되는 예제인 <sup>5</sup>에서는 새로운 카펫 세척제를 마케팅하려는 한 회사가 패키지 계획, 브랜드 이름, 값, 보증 유무 및 환불 유무라는 다섯 가지 요인이 고객 선호도에 어떤 영향을 미치는지 조사합니다. 패키지 계획에는 세 가지 요인 수준이 있고 그 각각은 세척용 솔의 위치에서 차이가 납니다. 그밖에도 세 가지 브랜드 이름(*K2R*, *Glory*, *Bissell*)이 있으며, 값 수준도 세 가지이고, 마지막 두 요인에 대해서는 각각 두 가지 수준(예 또는 아니오)이 있습니다. 이들 요인으로 정의된 22개의 프로파일에 대해 10명의 고객이 순위를 매겼습니다. 변수 선호도에는 각 프로파일의 평균 순위가 포함되어 있습니다. 낮은 순위는 선호도가 높음을 의미합니다. 이 변수는 각 프로파일에 대한 선호도의 전반적인 측도를 반영합니다.
- **carpet\_prefs.sav.** 이 데이터 파일은 *carpet.sav*에 설명된 것과 동일한 예제를 기반으로 하지만 10명의 고객 각각으로부터 수집한 실제 순위를 포함합니다. 고객들에게 가장 선호하는 제품부터 가장 덜 선호하는 제품에 이르기까지 22개 제품 프로파일에 순위를 매기도록 요청하였습니다. 변수 *PREF1*부터 *PREF22*까지는 *carpet\_plan.sav*에 정의된 관련 프로파일의 식별자를 포함합니다.
- **catalog.sav.** 이 데이터 파일에는 카탈로그 회사에서 판매한 세 가지 제품의 가설적인 월별 판매 수치가 들어 있습니다. 가능한 5 가지 예측변수에 대한 데이터도 포함되어 있습니다.
- **catalog\_seasfac.sav.** 이 데이터 파일은 부수적인 날짜 변수와 함께 계절 분해 프로시저에서 계산한 계절 요인 그룹을 추가한 것을 제외하고는 *catalog.sav*와 동일합니다.
- **cellular.sav.** 이것은 서비스 제공 사업자 이탈을 줄이려는 이동전화 회사의 시도와 관련이 있는 가설 데이터 파일입니다. 서비스 제공 사업자 이탈 성향 스코어가 0 ~ 100 범위로 계정에 적용됩니다. 스코어가 50이상인 계정은 서비스 제공업체를 변경할 것을 기대할 수 있습니다.
- **ceramics.sav.** 이것은 새로운 프리미엄급 합금이 표준 합금보다 내열성이 훨씬 큰지 여부를 판단하기 위한 제조업체의 시도와 관련이 있는 가설 데이터 파일입니다. 각 케이스는 합금 중 한 가지에 대한 별도의 검정을 나타냅니다. 베어링이 파손되는 온도를 기록합니다.
- **cereal.sav.** 이것은 880명의 사람들을 대상으로 연령, 성별, 결혼 상태 및 활동적인 라이프스타일을 갖고 있는지 여부(1주에 적어도 두 번은 운동을 하고 있는지)와 함께 아침 식사 메뉴 선호도를 조사한 여론조사와 관련이 있는 가설 데이터 파일입니다. 각 케이스는 개별 반응자를 나타냅니다.

4. McCullagh, P., and J. A. Nelder. 1989. *Generalized Linear Models*, 2nd ed. London: Chapman & Hall.

5. Green, P. E., and Y. Wind. 1973. *Multiaattribute decisions in marketing: A measurement approach*. Hinsdale, Ill.: Dryden Press.

- **clothing\_defects.sav.** 이것은 의류 공장에서 품질 관리 프로세스와 관련이 있는 가설 데이터 파일입니다. 공장에서 생산된 각 로트로부터 검사자들은 의복 표본을 선택하고 채택이 불가능한 의복 개수를 셉니다.
- **coffee.sav.** 이 데이터 파일은 여섯 가지 아이스커피 브랜드에 대해 인지된 이미지에 관한 데이터 파일입니다<sup>6</sup>. 23개의 아이스커피 이미지 속성 각각에 대해 해당 속성으로 설명할 수 있는 브랜드를 모두 선택하도록 사람들에게 요청했습니다. 브랜드 이름이 노출되지 않도록 여섯 개의 브랜드에는 AA, BB, CC, DD, EE 및 FF라고 이름을 붙였습니다.
- **contacts.sav.** 이것은 기업 컴퓨터 판매 대리점 그룹의 담당자 목록과 관련이 있는 가설 데이터 파일입니다. 각 담당자는 이들이 일하는 회사의 부서 및 그들의 회사내 지위별로 범주가 나누어졌습니다. 또한 마지막 판매량, 최종 판매 이후 시간 및 담당자 회사의 규모를 기록하였습니다.
- **creditpromo.sav.** 이것은 최근 신용카드 판촉의 효과를 평가하기 위한 백화점의 시도와 관련이 있는 가설 데이터 파일입니다. 이러한 목적을 위하여 500명의 카드 소지자를 무작위로 선택하였습니다. 이중 절반은 다음 세 달 동안 구매에 대한 이자율을 줄여 주겠다는 판촉 광고물을 받았습니다. 절반은 표준적인 시즌 광고물을 받았습니다.
- **customer\_dbase.sav.** 이것은 데이터 웨어하우스의 정보를 사용하여 답변할 가능성이 가장 높은 고객들에게 특가 제공을 지원하려는 회사의 시도와 관련 있는 가설 데이터 파일입니다. 고객 층의 서브세트를 무작위로 선택하여 특가 제공을 지원하고 그들의 반응을 기록하였습니다.
- **customer\_information.sav.** 이름 및 주소와 같은 고객 메일링 정보가 들어 있는 가설 데이터 파일입니다.
- **customer\_subset.sav.** *customer\_dbase.sav*의 80개 케이스의 서브세트입니다.
- **debate.sav.** 이것은 정치적 토론의 참가자들을 대상으로 한 토론 전후 설문조사에서 대응반응과 관련 있는 가설 데이터 파일입니다. 각 케이스는 개별 반응자에 해당합니다.
- **debate\_aggregate.sav.** 이것은 *debate.sav*의 반응을 통합하는 가설 데이터 파일입니다. 각 케이스는 토론 전후 선호도에 대한 교차 분류에 해당합니다.
- **demo.sav.** 이것은 매달 메일을 발송할 목적으로 구매 고객 데이터베이스와 관련 있는 가설 데이터 파일입니다. 다양한 인구 통계학 정보와 함께 고객이 제공에 반응했는지 여부가 기록됩니다.
- **demo\_cs\_1.sav.** 이것은 설문조사 정보 데이터베이스를 컴파일하려는 회사의 시도중 첫 단계와 관련 있는 가설 데이터 파일입니다. 각 케이스는 서로 다른 도시에 해당하며 지역, 지방, 구 및 도시 식별이 기록됩니다.
- **demo\_cs\_2.sav.** 이것은 설문조사 정보 데이터베이스를 컴파일하려는 회사의 시도중 두 번째 단계와 관련 있는 가설 데이터 파일입니다. 각 케이스는 첫 번째 단계에서 선정된 도시의 서로 다른 가구 단위에 해당하며 지역, 지방, 구, 도시, 구역 및 단위 식별이 기록됩니다. 계획 중 처음 두 단계의 표본 추출 정보도 포함됩니다.

---

6. Kennedy, R., C. Riquier, and B. Sharp. 1996. Practical applications of correspondence analysis to categorical data in market research. *Journal of Targeting, Measurement, and Analysis for Marketing*, 5, 56-70.

- **demo\_cs.sav.** 이것은 복잡한 표본추출 계획을 사용하여 수집한 설문조사 정보를 포함하고 있는 가설 데이터 파일입니다. 각 케이스는 서로 다른 가구 단위에 해당하며 다양한 인구 통계학적 정보와 표본 추출 정보가 기록됩니다.
- **diabetes\_costs.sav.** 보험 회사가 당뇨병이 있는 보험 계약자에 대해 유지보수하는 정보가 포함된 가설 데이터 파일입니다. 각 케이스는 서로 다른 보험 가입자에 해당합니다.
- **dietstudy.sav.** 이것은 "스틸만 다이어트"의 연구 결과를 포함하는 가설 데이터 파일입니다.<sup>7</sup> 각 케이스는 별도의 주제에 해당하며 파운드 단위로 다이어트 전후 체중을 기록하고 나무글리세이드 레벨을 mg/100 ml 단위로 기록합니다.
- **dmdata.sav.** 다이렉트 마케팅 회사의 인구 통계학 및 구매 정보가 들어 있는 가설 데이터 파일입니다. *dmdata2.sav*에는 검정 메일링을 수신한 연락처 서브세트에 대한 정보가 들어 있으며, *dmdata3.sav*에는 검정 메일을 수신하지 않은 나머지 연락처에 대한 정보가 들어 있습니다.
- **dvdplayer.sav.** 이것은 새 DVD 플레이어 개발과 관련 있는 가설 데이터 파일입니다. 마케팅 팀은 프로토타입을 사용하여 집중 그룹 데이터를 수집하였습니다. 각 케이스는 설문조사에 응한 각 사용자에게 해당하며 이들에 관한 몇 가지 인구 통계학적 정보와 프로토타입에 관한 양케이트 반응을 기록합니다.
- **german\_credit.sav.** 이 데이터 파일은 Irvine의 캘리포니아 대학<sup>8</sup> Repository of Machine Learning 데이터베이스의 "독일인의 신용(German credit)" 데이터 세트에서 추출했습니다.
- **grocery\_1month.sav.** 이것은 매주 "종합" 구매에 대한 *grocery\_coupons.sav* 데이터 파일내의 가설 데이터 파일입니다. 따라서 각 케이스는 각 고객에 해당합니다. 결과적으로 매주 변하는 변수중 일부는 사라지며 이제 기록된 소비 금액은 4주의 연구기간중 소비한 금액의 합계가 됩니다.
- **grocery\_coupons.sav.** 이것은 고객의 구매 습관에 관심 있는 식료품 체인점에서 수집한 설문조사 데이터를 포함한 가설 데이터 파일입니다. 각 고객을 4주 동안 추적하며 각 케이스는 각각의 고객-주에 해당합니다. 고객이 그 주 동안 식료품 상점에서 소비한 금액을 포함하여 어디서 어떻게 쇼핑했는지에 관한 정보를 기록합니다.
- **guttman.sav.** Bell<sup>9</sup>은 가능한 사회 그룹을 보여 주는 표를 제안했습니다. Guttman<sup>10</sup>은 이 표의 일부를 사용했습니다. 그 중 사회적 상호작용, 그룹 소속감, 멤버의 친밀감, 관계의 형식성 등을 설명하는 다섯 개의 변수는 일곱 개의 이론적인 사회 그룹과 교차합니다. 이 일곱 개의 그룹에는 군중(예: 축구 시합 관중), 청중(예: 극장 관객 또는 강연회 방청객), 대중(예: 신문 구독자 또는 TV 시청자), 폭도(군중과 비슷하지만 상호작용이 훨씬 더 강한 그룹), 일차그룹(친밀한 관계), 이차그룹(자발적 관계), 현대 공동체(강한 친밀감과 특별한 서비스에 대한 필요의 결과로 발생한 느슨한 연합)가 있습니다.

7. Rickman, R., N. Mitchell, J. Dingman, and J. E. Dalen. 1974. Changes in serum cholesterol during the Stillman Diet. *Journal of the American Medical Association*, 228:, 54-58.

8. Blake, C. L., and C. J. Merz. 1998. "UCI Repository of machine learning databases." Available at <http://www.ics.uci.edu/~mllearn/MLRepository.html>.

9. Bell, E. H. 1961. *Social foundations of human behavior: Introduction to the study of sociology*. New York: Harper & Row.

10. Guttman, L. 1968. A general nonmetric technique for finding the smallest coordinate space for configurations of points. *Psychometrika*, 33, 469-506.

- **health\_funding.sav.** 이것은 의료보험 기금(모집단 100명당 금액), 발병율(모집단 10000명당 비율), 의료보험 제공업체 방문자(모집단 10000명당 비율)에 대한 데이터를 포함하는 가설 데이터 파일입니다. 각 케이스는 서로 다른 도시를 나타냅니다.
- **hivassay.sav.** 이것은 HIV 감염 여부를 발견하기 위한 신속한 검사 분석표를 개발하기 위한 제약 실험실의 시도와 관련 있는 가설 데이터 파일입니다. 검사 분석표 결과는 적색의 8단계 농도로 표시됩니다. 농도가 짙을수록 감염 가능성이 더 높음을 나타냅니다. 실험실에서 2000명의 혈액 표본에 대해 시험한 결과 이중 절반이 HIV에 감염되었으며 절반은 깨끗했습니다.
- **hourlywagedata.sav.** 이것은 다양한 경력을 갖춘 진료실 및 병원의 간호사의 시간당 급여에 관한 가설 데이터 파일입니다.
- **insurance\_claims.sav.** 의심스럽거나, 사기성이 있는 청구를 거부하는 데 사용할 모형을 생성하려는 보험 회사를 위한 가설 데이터입니다. 각 케이스는 개별 청구를 나타냅니다.
- **insure.sav.** 이것은 고객이 10년만기 생명보험 계약에서 보험료를 청구할지 여부를 나타내는 위험 요인을 연구하는 보험 회사에 관한 가설 데이터 파일입니다. 데이터 파일의 각 케이스는 대응하는 계약을 나타내며, 그 중 한 계약은 연령과 성별에 따라 보험료 청구를 기록하였으며 다른 한 계약은 기록하지 않았습니다.
- **judges.sav.** 300개 체조 연기에 대해 훈련된 심판(+ 1명의 팬)이 매긴 스코어에 관한 가설 데이터 파일입니다. 각 행은 각각의 연기를 나타내며, 심판은 동일한 연기를 보았습니다.
- **kinship\_dat.sav.** Rosenberg와 Kim<sup>11</sup>은 친족 관계를 나타내는 15개의 단어(고모, 형제, 사촌, 딸, 아버지, 손녀, 할아버지, 할머니, 손자, 어머니, 조카, 조카딸, 누이, 아들, 삼촌)에 대한 조사를 착수했습니다. 조사자는 네 개의 대학생 그룹(여학생 두 그룹, 남학생 두 그룹)에 부탁하여 유사성을 기준으로 이들 단어를 정렬하도록 요청했습니다. 두 그룹(여학생 한 그룹, 남학생 한 그룹)에는 단어를 두 차례에 걸쳐 정렬하되 두 번째 정렬 기준은 첫 번째 경우와 다른 기준을 사용하도록 했습니다. 이런 식으로 6개 "소스"의 총계를 얻었습니다. 각 소스는 15 x 15 근접 행렬에 대응합니다. 이 행렬의 셀은 소스에 포함된 사람 수에서 개체가 해당 소스에서 함께 분할된 횟수를 뺀 값과 같습니다.
- **kinship\_ini.sav.** 이 데이터 파일에는 *kinship\_dat.sav*에 대한 3차원적 솔루션의 초기 설정이 들어 있습니다.
- **kinship\_var.sav.** 이 데이터 파일에는 *kinship\_dat.sav*의 해법 차수를 해석하는 데 사용될 수 있는 독립변수 *gender*(성별), *gener*(세대) 및 *degree* (촌수)가 포함되어 있습니다. 특히, 이들 독립변수는 이러한 변수의 선형 조합으로 해법 공간을 제한하는 데 사용할 수 있습니다.
- **marketvalues.sav.** 이 데이터 파일은 1999 – 2000년 동안 Algonquin, Ill. 지역의 신규 주택 개발의 판매량에 관한 데이터 파일입니다. 이 판매량은 공식 레코드에 남아 있는 사항입니다.
- **nhis2000\_subset.sav.** NHIS(National Health Interview Survey)는 미국 일반 시민을 대상으로 한 대규모의 인구 기반 설문조사입니다. 인터뷰는 국가 전역에서 뽑은 대표적인 표본 세대를 직접 대면하는 방식으로 진행합니다. 각 세대 멤버와의 인터뷰를 통해 각자의 생활 습관 및 건강 상태에

---

11. Rosenberg, S., and M. P. Kim. 1975. The method of sorting as a data-gathering procedure in multivariate research. *Multivariate Behavioral Research*, 10, 489-502.

대한 관측값과 인구 통계학적 정보를 수집합니다. 이 데이터 파일은 2000년 설문조사의 정보 서브 세트를 포함합니다. 미국 국립 건강 통계 센터 2000년 미국 국민 건강 설문조사, 공용 데이터 파일 및 문서자료. [ftp://ftp.cdc.gov/pub/Health\\_Statistics/NCHS/Datasets/NHIS/2000/](ftp://ftp.cdc.gov/pub/Health_Statistics/NCHS/Datasets/NHIS/2000/). 2003년 액세스됨.

- **ozone.sav.** 이 데이터에는 남은 변수로부터 오존 농도를 예측하기 위해 6가지 기상학적 변수에 관한 330개 관측값이 포함되어 있습니다. 이전의 조사<sup>12</sup>, 무엇보다도<sup>13</sup>에서 이들 변수 간에 비선형성을 발견했으며 이는 표준 회귀분석 방법을 사용하는 데 걸림돌이 되고 있습니다.
- **pain\_medication.sav.** 이것은 만성 관절염 통증 치료를 위한 항염증제 임상 시험 결과를 포함한 가설 데이터 파일입니다. 특히 관심을 받는 것은 약물이 효능을 발휘하는 데 소요되는 시간이며 기존의 약제와 비교하는 방법입니다.
- **patient\_los.sav.** 이것은 심근경색(MI, 또는 심장마비)이 의심되어 병원에 내원한 환자들을 대상으로 한 치료 레코드가 포함된 가설 데이터 파일입니다. 각 케이스는 각 환자에 해당하며 환자의 입원에 관한 수많은 변수를 기록합니다.
- **patlos\_sample.sav.** 이것은 심근경색(MI, 또는 심장마비)을 치료하는 동안 혈전 용해제를 투여받은 환자 표본에 대한 치료 레코드가 포함된 가설 데이터 파일입니다. 각 케이스는 각 환자에 해당하며 환자의 입원에 관한 수많은 변수를 기록합니다.
- **poll\_cs.sav.** 이것은 입법 전 법안에 대한 공공 지지도를 판단하려는 여론 설문조사원의 시도에 관한 가설 데이터 파일입니다. 케이스는 등록된 유권자에 해당합니다. 각 케이스는 유권자가 살고 있는 군, 읍, 지구를 기록합니다.
- **poll\_cs\_sample.sav.** 이 가설 데이터 파일에는 *poll\_cs.sav*에 등록된 유권자 표본이 들어 있습니다. 표본은 *poll\_csplan* 계획 파일에 지정된 계획에 따라 선택되었으며 포함 확률 및 표본 가중치를 기록합니다. 그러나 표본추출 계획은 PPS(크기 비례) 방법을 사용하기 때문에 결합 선택 확률을 포함한 파일도 있음을 유의하십시오(*poll\_jointprob.sav*). 유권자 인구 통계학 정보에 해당하는 추가 변수 및 제안된 법안에 관한 그들의 의견을 수집하였으며 표본을 추출한 후 데이터 파일에 추가하였습니다.
- **property\_assess.sav.** 이것은 제한된 자원에 대해 현재까지의 자산 가치를 평가하려는 카운티 평가자의 시도와 관련 있는 가설 데이터 파일입니다. 케이스는 작년 카운티에서 판매된 자산에 해당합니다. 데이터 파일의 각 케이스는 자산이 있는 군/구, 자산을 최종 방문한 평가자, 평가 후 경과 시간, 그 당시 감정 결과, 자산에 대한 판매 가치를 기록합니다.
- **property\_assess\_cs.sav.** 이것은 제한된 자원에 대해 현재까지의 자산 가치를 평가하려는 주(州) 평가자의 시도와 관련 있는 가설 데이터 파일입니다. 케이스는 주의 자산에 해당합니다. 데이터 파일의 각 케이스는 자산이 있는 군, 읍, 지구와 최종 평가 이후 경과된 시간, 그 당시 감정 결과를 기록합니다.

---

12. Breiman, L., and J. H. Friedman. 1985. Estimating optimal transformations for multiple regression and correlation. *Journal of the American Statistical Association*, 80, 580-598.

13. Hastie, T., and R. Tibshirani. 1990. *Generalized additive models*. London: Chapman and Hall.

- **property\_assess\_cs\_sample.sav.** 이 가설 데이터 파일에는 *property\_assess\_cs.sav*에 등록된 자산 표본이 들어 있습니다. 표본은 *property\_assess.csplan* 계획 파일에 지정된 계획에 따라 선택되었으며 포함 확률 및 표본 가중치를 기록합니다. 추가 변수 현재 값이 수집되었으며 표본을 선택한 후 데이터 파일에 추가되었습니다.
- **recidivism.sav.** 이것은 관할 영역에서 전과자 비율을 파악하려는 정부 법률 집행 기관의 시도와 관련 있는 가설 데이터 파일입니다. 각 케이스는 이전의 범죄자에 해당하며 그들의 인구 통계학 정보, 첫 번째 범죄 상세 내용, 첫 번째 구속 후 2년내에 범죄가 재발한 경우 두 번째 구속까지 경과된 시간을 기록합니다.
- **recidivism\_cs\_sample.sav.** 이것은 관할 영역에서 전과자 비율을 파악하려는 정부 법률 집행 기관의 시도와 관련 있는 가설 데이터 파일입니다. 각 케이스는 2003년 6월 동안 처음 구속된 후 출감한 이전의 범죄자에 해당하며 그들의 인구 통계학 정보, 첫 번째 범죄 상세 내용, 2006년 6월 말까지 범죄가 재발한 경우 두 번째 구속까지 경과된 시간을 기록합니다. 범죄자들은 *recidivism\_cs.csplan*에 지정된 표본추출 계획에 따라 표본 추출된 부서에서 선택되었습니다. 왜냐하면 표본추출 계획은 PPS(크기 비례) 방법을 사용하므로 결합 선택 확률을 포함한 파일(*recidivism\_cs\_jointprob.sav*)도 있기 때문입니다.
- **rfm\_transactions.sav.** 구매 데이터, 구매한 항목 및 각 트랜잭션 금액이 포함된 구매 트랜잭션 데이터가 있는 가설 데이터 파일입니다.
- **salesperformance.sav.** 이것은 두 가지 새 영업 교육 강의 평가와 관련 있는 가설 데이터 파일입니다. 60명의 직원들을 세 그룹으로 구분하였고 모두에게 표준 교육을 받게 하였습니다. 또한 그룹 2에는 기술 교육을, 그룹 3에는 실무 자습서를 제공합니다. 교육 강의를 끝나면 각 피고용자를 검정하여 스코어를 기록합니다. 데이터 파일의 각 케이스는 개별 훈련생을 나타내며 할당된 그룹과 시험 성적을 기록합니다.
- **satisf.sav.** 4군데 상점에 입점한 유통업체에서 실시간 고객 만족도 설문조사에 관한 가설 데이터 파일입니다. 582명의 고객 모두에 대해 설문조사를 실시하였으며 각 케이스는 한 명의 고객으로부터 받은 반응을 나타냅니다.
- **screws.sav.** 이 데이터 파일에는 나사, 볼트, 너트 및 못의 특성에 대한 정보가 포함되어 있습니다.
- **shampoo\_ph.sav.** 이것은 모발 관리 제품 공장의 품질 관리 프로세스와 관련이 있는 가설 데이터 파일입니다. 정기적인 시간 간격으로 6벌의 서로 다른 제품을 측정하여 pH를 기록하였습니다. 대상 범위는 4.5 - 5.5입니다.
- **ships.sav.** 파도에 의한 화물선박 손상에 관해 <sup>15</sup> 이외에서 제시 및 분석된 데이터 세트입니다. 사고 개수는 선박의 종류, 건조 기간, 서비스 기간을 가정한 포아송 비율에서 발생하는 것으로 모델링할 수 있습니다. 요인의 교차 분류에 의해 형성된 각 표 셀의 총 서비스 개월수는 위험에 대한 노출 값을 제공합니다.

14

14. Hartigan, J. A. 1975. *Clustering algorithms*. New York: John Wiley and Sons.

15. McCullagh, P., and J. A. Nelder. 1989. *Generalized Linear Models*, 2nd ed. London: Chapman & Hall.

- **site.sav.** 이것은 사업 확장을 위해 새 부지를 선택하려는 회사의 시도와 관련 있는 가설 데이터 파일입니다. 이 회사는 두 명의 컨설턴트를 고용하여 부지에 대한 전망을 따로따로 평가하게 하였고 확장 보고서에서는 각 부지를 "우수", "양호" 또는 "불량"으로 요약하였습니다.
- **smokers.sav.** 이 데이터 파일은 1998년 전국 마약 남용에 관한 가정 설문조사에서 추출한 것으로 미국 가정의 확률 표본입니다. (<http://dx.doi.org/10.3886/ICPSR02934>) 따라서 이 데이터 파일 분석의 첫 단계는 인구 동향을 반영하는 데이터에 가중치를 더하는 작업이 되어야 합니다.
- **stocks.sav** 이 가설 데이터 파일에는 일 년 동안의 주가와 거래량이 들어 있습니다.
- **stroke\_clean.sav.** 이것은 Statistics Base Edition의 절차를 사용하여 삭제한 의료 데이터베이스의 상태를 포함한 가설 데이터 파일입니다.
- **stroke\_invalid.sav.** 이 가설 데이터 파일은 의료 데이터베이스의 초기 상태와 여러 데이터 입력 오류를 포함합니다.
- **stroke\_survival.** 이 가설 데이터 파일은 허혈성 뇌졸중 발생 후 수많은 도전과제에 직면하고 있는 재활 프로그램을 종료한 환자들의 생존 시간과 관련 있습니다. 뇌졸중, 심장마비의 발생, 허혈성 뇌졸중 또는 출혈성 뇌졸중으로 구분되며 이벤트 발생 시간이 기록됩니다. 표본은 현재까지 살아남은 사람들에 대한 데이터입니다. 왜냐하면 뇌졸중 후 시행한 재활 프로그램 마지막에 생존한 환자만 포함하기 때문입니다.
- **stroke\_valid.sav.** 이것은 데이터 검증 절차를 사용하여 값을 점검한 후 의료 데이터베이스의 상태를 포함한 가설 데이터 파일입니다. 여기에는 여전히 잠재적인 이상항목 케이스가 포함됩니다.
- **survey\_sample.sav.** 이 데이터 파일은 인구 통계학 데이터 및 다양한 태도 측도 등 설문조사 데이터가 포함됩니다. 이 데이터 파일은 1998 NORC General Social Survey의 변수 서브셋을 기준으로 하지만, 설명을 목적으로 일부 데이터 값이 수정되었으며 추가 가상 변수가 추가되었습니다.
- **tcm\_kpi.sav.** 이것은 비즈니스의 매주 핵심성과지표(KPI) 값이 포함된 가설 데이터 파일입니다. 여기에는 또한 동일한 기간 동안 제어 가능한 다수의 메트릭에 대한 매주 데이터가 포함됩니다.
- **tcm\_kpi\_upd.sav.** 이 데이터 파일은 *tcm\_kpi.sav*와 동일하나 추가 4주에 대한 데이터가 들어 있습니다.
- **telco.sav** 이것은 고객층에서 서비스 제공 사업자 이탈을 줄이려는 통신 회사의 시도와 관련이 있는 가설 데이터 파일입니다. 각 케이스는 개별 고객에 해당하며 다양한 인구 통계학 정보와 서비스 사용 정보가 기록됩니다.
- **telco\_extra.sav.** 이 데이터 파일은 *telco.sav* 데이터 파일과 유사하지만 "tenure"(보유 기간) 및 로그 변환된 고객 소비 변수는 제거되었고 표준화된 로그 변환된 고객 소비 변수로 대체되었습니다.
- **telco\_missing.sav.** 이 데이터 파일은 *telco.sav* 데이터 파일 서브셋이지만 인구 통계학적 정보 데이터 값 중 일부는 결측값으로 대체되었습니다.
- **testmarket.sav.** 이 가설 데이터 파일은 메뉴에 새로운 품목을 추가하려는 패스트푸드 체인점의 계획과 관련이 있습니다. 신제품을 판촉하는 세 가지 가능한 캠페인이 있습니다. 따라서 무작위로 선택된 여러 시장 위치에 새 품목이 도입됩니다. 각 위치에 서로 다른 판촉을 사용하며 처음 4주 동안 새 품목의 주간 판매량을 기록합니다. 각 케이스는 별도의 위치-주에 해당합니다.

- **testmarket\_1month.sav.** 이것은 매주 "종합" 판매량에 대한 *testmarket.sav* 데이터 파일내의 가설 데이터 파일입니다. 따라서 각 케이스는 각 위치에 해당합니다. 결과적으로 매주 변하는 변수중 일부는 사라지며 이제 기록된 판매량은 4주의 연구기간중 판매량의 합계가 됩니다.
- **tree\_car.sav.** 이것은 인구 통계학 및 차량 구매 가격 데이터를 포함한 가설 데이터 파일입니다.
- **tree\_credit.sav.** 이것은 인구 통계학 및 은행 대부 이력을 포함한 가설 데이터 파일입니다.
- **tree\_missing\_data.sav** 이것은 수많은 결측값을 갖고 있는 인구 통계학 데이터 및 은행 대부 이력 데이터를 포함한 가설 데이터 파일입니다.
- **tree\_score\_car.sav.** 이것은 인구 통계학 및 차량 구매 가격 데이터를 포함한 가설 데이터 파일입니다.
- **tree\_textdata.sav.** 측정 수준 및 값 레이블을 할당하기 전 주로 기본값 변수 상태를 보여주기 위한 목적으로 두 변수만을 포함한 단순 데이터 파일입니다.
- **tv-survey.sav.** 이것은 성공적인 프로그램 운영을 확장할지 여부를 고려하기 위해 텔레비전 스튜디오에서 실시한 여론설문조사에 관한 가설 데이터 파일입니다. 906명의 반응자에게 다양한 조건 하에서 프로그램을 시청할지 여부에 대해 질문하였습니다. 각 행은 개별 반응자를 나타내며 각 열은 개별 조건을 나타냅니다.
- **ulcer\_recurrence.sav.** 이 파일에는 궤양의 재발 방지를 위한 두 가지 치료법의 효능을 비교하도록 계획된 연구의 정보 일부가 들어 있습니다. 이것은 구간 중도 절단 자료(interval-censored data)의 좋은 예를 제공하며 <sup>16</sup> 이외의 장소에서 제시 및 분석되었습니다.
- **ulcer\_recurrence\_recoded.sav.** 이 파일은 *ulcer\_recurrence.sav*의 정보를 재정리하므로 단순히 연구가 끝난 후의 이벤트 확률이 아닌 연구의 각 구간에 대한 이벤트 확률을 모델링할 수 있습니다. 이것은 <sup>17</sup> 이외의 장소에서 제시 및 분석되었습니다.
- **verd1985.sav.** 이 데이터 파일은 설문조사 <sup>18</sup>과 관련 있습니다. 15명의 사람들이 8개의 변수에 대해 반응한 내용을 기록했습니다. 관심 변수를 세 가지 세트로 구분하였습니다. 변수 세트 1에는 연령과 결혼 상태가 포함되고, 변수 세트 2에는 애완동물과 신문이 포함되고, 변수 세트 3에는 음악과 거주가 포함됩니다. 애완동물은 다중 명목으로 척도화되고 연령은 순서로 척도화됩니다. 다른 모든 변수는 단일 명목으로 척도화됩니다.
- **virus.sav.** 이것은 바이러스가 네트워크에 미치는 영향을 판단하려는 인터넷 서비스 공급업체(ISP)의 시도와 관련 있는 가설 데이터 파일입니다. 발견 시점에서 위협이 제거된 시점까지의 시간 동안 자사의 네트워크에서 감염된 이메일 트래픽의 대략적인 퍼센트를 추적했습니다.

---

16. Collett, D. 2003. *Modelling survival data in medical research*, 2 ed. Boca Raton: Chapman & Hall/CRC.

17. Collett, D. 2003. *Modelling survival data in medical research*, 2 ed. Boca Raton: Chapman & Hall/CRC.

18. Verdegaal, R. 1985. *Meer sets analyse voor kwalitatieve gegevens (in Dutch)*. Leiden: Department of Data Theory, University of Leiden.

- **wheeze\_steubenville.sav.** 이것은 대기 오염이 자녀들의 건강에 미치는 영향에 대한 종단적 연구의 서브세트입니다.<sup>19</sup> 데이터에는 오하이오주 Steubenville의 7~10세 어린이들을 대상으로 연구 첫째 동안은 어머니가 흡연자인지 여부에 관한 레코드를 포함, 호흡기 천명에 대한 이원적 반복 측도 결과가 들어 있습니다.
- **workprog.sav.** 이것은 불리한 조건을 가진 사람들을 더 나은 직업으로 배치하려는 정부 연구 프로그램에 관한 가설 데이터 파일입니다. 잠재적인 프로그램 참가자 표본을 추적하였습니다. 이들 중 일부는 프로그램 영입을 위하여 무작위로 선택된 반면 다른 이들은 그렇지 않습니다. 각 케이스는 개별 프로그램 참가자를 나타냅니다.
- **worldsales.sav** 이 가설 데이터 파일에는 대륙 및 제품별 판매 수입이 들어 있습니다.

---

19. Ware, J. H., D. W. Dockery, A. Spiro III, F. E. Speizer, and B. G. Ferris Jr.. 1984. Passive smoking, gas cooking, and respiratory health of children living in six cities. *American Review of Respiratory Diseases*, 129, 366-374.



---

## 주의사항

이 정보는 미국에서 제공되는 제품 및 서비스용으로 작성된 것입니다. 본 자료는 다른 언어로도 제공될 수 있습니다. 그러나 자료에 접근하기 위해서는 해당 언어로 된 제품 또는 제품 버전의 사본이 필요할 수 있습니다.

IBM은 다른 국가에서 이 책에 기술된 제품, 서비스 또는 기능을 제공하지 않을 수도 있습니다. 현재 사용할 수 있는 제품 및 서비스에 대한 정보는 한국 IBM 담당자에게 문의하십시오. 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.

07326

서울특별시 영등포구  
국제금융로 10, 31FC  
한국 아이.비.엠 주식회사  
고객만족센터

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

*Intellectual Property Licensing*

*Legal and Intellectual Property Law*

*IBM Japan Ltd.*

*19-21, Nihonbashi-Hakozakicho, Chuo-ku*

*Tokyo 103-8510, Japan*

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

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

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

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

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

07326

서울특별시 영등포구

국제금융로 10, 31FC

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

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

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

IBM의 향후 제시 방향 또는 의도에 관한 언급은 특별한 통지 없이 변경 내지 철회될 수 있으며, 단순히 목표만을 의미합니다.

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

저작권 라이선스:

이 정보에는 여러 운영 플랫폼에서의 프로그래밍 기법을 보여주는 원어로 된 샘플 응용프로그램이 들어 있습니다. 귀하는 이러한 샘플 프로그램의 작성 기준이 된 운영 플랫폼의 애플리케이션 프로그래밍 인터페이스(API)에 부합하는 애플리케이션을 개발, 사용, 판매 또는 배포할 목적으로 IBM에 추가 비

용을 지불하지 않고 이들 샘플 프로그램을 어떠한 형태로든 복사, 수정 및 배포할 수 있습니다. 이러한 샘플 프로그램은 모든 조건하에서 완전히 테스트된 것은 아닙니다. 따라서 IBM은 이러한 프로그램의 신뢰성, 서비스 가능성 또는 기능을 보증하거나 진술하지 않습니다. 본 샘플 프로그램은 일체의 보증 없이 "현상태대로" 제공됩니다. IBM은 귀하의 샘플 프로그램 사용과 관련되는 손해에 대해 책임을 지지 않습니다.

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

© IBM 2019. 이 코드의 일부는 IBM Corp.의 샘플 프로그램에서 파생됩니다.

© Copyright IBM Corp. 1989 - 2019. All rights reserved.

---

## 상표

IBM, IBM 로고 및 [ibm.com](http://ibm.com)은 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](http://www.ibm.com/legal/copytrade.shtml))의 "저작권 및 상표 정보"에 있습니다.

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

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

Linux는 미국 또는 기타 국가에서 사용되는 Linus Torvalds의 등록상표입니다.

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

UNIX는 미국 및 기타 국가에서 사용되는 The Open Group의 등록상표입니다.

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



## 색인

### [가]

값 레이블  
  뷰어에서 표시 제어 25  
  숫자변수 25  
  할당 25  
값 코딩변경 73  
개수  
  개수 표 29  
결과 내보내기  
  Excel 57  
  HTML 65  
  PowerPoint 57  
  Word 57  
결측값  
  비슷자 변수용 27  
  숫자변수 27  
  시스템 결측값 26  
구간 데이터 29  
구문 67  
구문 도움말 도구 68  
구문 붙여넣기  
  대화 상자 67  
구문 창  
  명령 붙여넣기 67  
  명령 실행 67, 69  
  색상 코딩 68  
  자동-완료 68  
  종단점 70  
  편집 명령 68  
구문 파일  
  열기 69  
그래프  
  그래프 작성 35  
  막대도표 35

### [나]

날짜 및 시간 마법사 78  
날짜 및 시간변수 78

### [다]

데이터 유형  
  변수 24  
데이터 입력 21, 23  
  비슷자 23  
  숫자 21  
데이터 편집기  
  비슷자 데이터 입력 23  
  숫자 데이터 입력 21  
데이터베이스 마법사 13  
데이터베이스 파일  
  읽기 13  
도표  
  도표 작성 35  
  막대도표 30, 35  
  히스토그램 32

### [라]

레이어  
  피벗 표에서 작성 45

### [마]

막대도표 30  
명목 데이터 29  
문자열 데이터  
  데이터 입력 23

### [바]

범주형 데이터 29  
  요약 측도 29  
변수 21  
  데이터 유형 24  
  레이블 23  
변수 레이블  
  작성 23  
변수 레이블 작성 23  
분할 파일 처리 84  
뷰어  
  결과 감추기 및 보이기 41  
  결과 이동 41

비율 데이터 29  
빈도표 29

### [사]

새로운 변수 계산 75  
순서 데이터 29  
숫자 데이터 21  
스프레드시트 파일  
  변수 이름 읽기 10  
  읽기 10  
시스템-결측값 26

### [아]

양적 데이터 29  
연속형 데이터 29  
요약 측도  
  범주형 데이터 29  
  척도변수 31  
이동  
  뷰어의 항목 41  
  피벗 표의 요소 43

### [자]

조건식 77  
질적 데이터 29

### [차]

척도 데이터 29  
척도변수  
  요약 측도 31  
측정 수준 29

### [카]

케이스  
  선택 86  
  정렬 83, 85  
케이스 서브세트  
  날짜 및 시간 기준 89  
  무작위 표본 88

케이스 서브세트 (계속)  
  선택 86  
  선택하지 않은 케이스 삭제 90  
  선택하지 않은 케이스 필터링 90  
  조건식 87  
  조건을 만족하는 케이스 87  
케이스 선택 86  
케이스 정렬 83

## [타]

텍스트 가져오기 마법사 17  
텍스트 데이터 파일  
  읽기 17

## [파]

표본 파일  
  위치 91  
표현식에 함수 77  
피벗 표  
  레이어 45  
  셀 데이터 유형 47  
  셀 형식 47  
  소수점 감추기 47  
  정의 액세스 42  
  편집 45  
  피벗 트레이 43  
  행과 열 감추기 46  
  행과 열 전치 43  
  형식 45  
피벗 표 편집 45  
피벗 표에서 행과 열 감추기 46  
피벗 표에서 행과 열 전치(이동) 43

## [하]

히스토그램 32

## E

Excel 파일  
  읽기 10  
Excel(Microsoft)  
  결과 내보내기 57

## H

HTML  
  결과 내보내기 65

## M

Microsoft Access 13

## P

PowerPoint(Microsoft)  
  결과 내보내기 57

## W

Word(Microsoft)  
  결과 내보내기 57



