

DBSTAT 에서 엑셀파일 사용

엑셀파일을 데이터베이스 파일로 사용하기 위해서는 몇 가지 규칙이 있습니다.

1. 변수명 설정

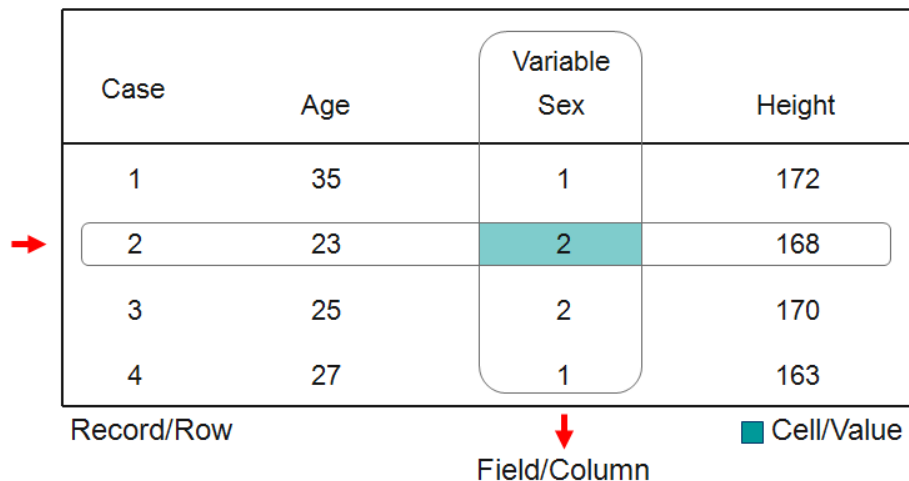
- 1) 변수명은 엑셀에서 첫 번째 행(row)에만 입력하여야 합니다.
- 2) 변수명은 인터넷에서 ID 설정과 동일합니다.
영문자로 시작하여야 하며 10 자리 수 이내에서 영문자, 숫자, _ (under bar)는 사용 가능
- 3) 한글, 특수문자(! @ # \$ % ^ & * () + - = , . < > / ?), 공백 사용 금지
- 4) 올바른 예: ID_NO, Age, Height, Year_2015, Department
- 5) 잘못된 예: ID NO(공백사용), 연령(한글사용), Wt%(특수문자) 2015_Year(숫자로 시작)

2. 각각의 변수(Variable)는 엑셀에서 열(column)에 해당합니다.

- 1) 변수의 형태에 따라 자료입력 - 숫자 행(필드)에는 숫자만, 문자 행(필드)에는 문자만 입력
- 2) 열과 열 사이에 빈 열이 있으면 안됩니다.

3. 각각의 레코드(Record)는 엑셀에서 행(row)에 해당합니다.

- 1) 하나의 레코드는 한 행에만 기록하여야 합니다.
- 2) 행과 행 사이에 빈 행이 있으면 안됩니다.



Case	Age	Variable Sex	Height
1	35	1	172
2	23	2	168
3	25	2	170
4	27	1	163

Record/Row

Field/Column

Cell/Value

4. 엑셀 쉬트(Sheet) 내 셀에 함수, 수식이나 포맷 사용 금지

- 1) 수식은 숫자로 바꾸어 주는 것이 좋습니다.
- 2) 엑셀에서 사용하는 특수한 포맷은 제거하는 것이 이상한 오류를 방지할 수 있습니다.

5. 날짜는 "YYYY-MM-DD" 형태로 입력

- 1) 엑셀에서 "2015-02-22" 형태로 입력하면 자동으로 날짜로 인식됩니다.

- 2) 달력 날짜를 사용하여 직접 통계 분석하지 않으므로 엑셀에서 날짜 계산을 하여 Month, Day 등 숫자 변수로 바꾸면 다른 통계패키지에서 불러들일 때에 사용하기 편리합니다.

다른 통계패키지에서 엑셀파일을 사용하려면 통계분석과 필요 없는 변수(행), Coding 에 대한 설명 등 메모를 삭제하여 정리하고 다른 이름으로 저장한 다음에 사용하는 방법이 오류를 줄이는 지름길입니다.

잘못된 변수명이 너무 많아 오류를 수정하기가 어려울 때에는 첫 번째 행을 삭제하고 입력된 자료를 첫 번째 행으로 하면 변수명 입력 오류를 없앨 수 있습니다. 이 경우에는 엑셀을 불러들이는 패키지에 따라서 변수명이 임의로 정해집니다.

dBSTAT 에서 Excel File 열기

1. 엑셀 파일에 여러 개의 쉬트(sheet)가 있으면 첫 번째 쉬트만 열 수 있습니다.

2. dBSTAT에서는 엑셀파일 첫 번째 행(row)을 변수명으로 불러들입니다.

- 1) 변수명이 한글인 경우 "Vn" (n: 변수 행) 형태로 자동 변경합니다.
- 2) 변수명이 특수문자인 경우 "_" (underbar)로 바꾸어 줍니다.
- 3) 첫 번째 행에 변수명이 없으면 변수명이 V1, V2, ..., Vn 으로 자동적으로 만들어집니다.

3. 숫자 변수(행)에 문자가 섞여 있으면 문자 ➡ 0 (zero)로 바꾸어 처리합니다.

4. 숫자 변수(행)에 공백(빈 칸, cell)은 0 (zero)로 처리합니다.

- 1) 숫자 변수에서 숫자 0 과 공백은 결측값(missing value)으로 처리됩니다.
- 2) 숫자 0 과 결측값을 구분하기 위해서는 공백에 '99', '-9' 등 사용불가 숫자를 입력합니다.
예) Height, Weight, Cholesterol: 0 (결측값) – 키, 체중, 콜레스테롤이 0 가 될 수 없다.
예) Age(year), Temperature: 연구에 따라서 다르다. 신생아 경우 연령(년) = 0, Temperature 가 체온이 아닌 기온이면 0 이 입력 가능한 수이므로 결측값을 '-9' 또는 '-999' 등으로 바꾼다.

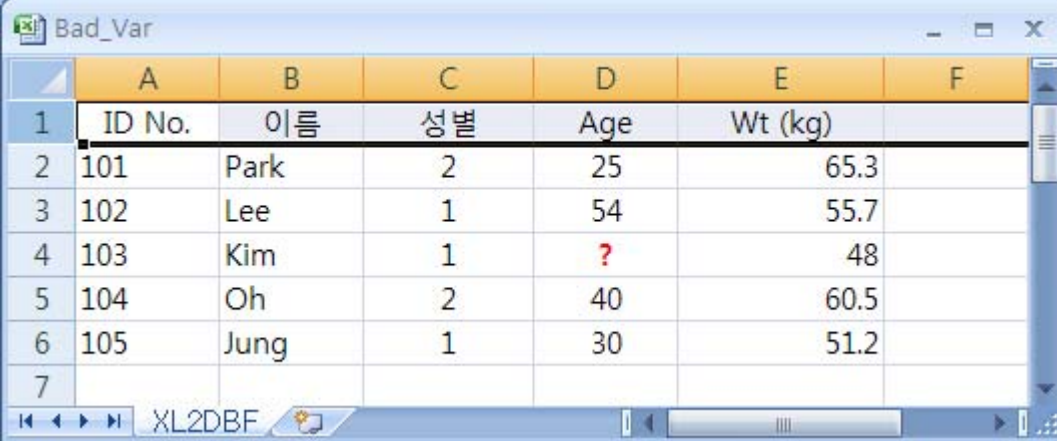
5. 엑셀에서 범주형인 숫자 변수(행)은 dBSTAT 에서 문자 필드로 바꾸어 주어야 합니다.

예) 엑셀에서 Sex(성별) 변수에 숫자 1, 2 로 입력 (Sex 1 = 여자, 2 = 남자)

Excel 파일 사용 예


1. 엑셀 파일: Bad Example

다음과 같은 엑셀 파일은 변수명에 한글, 특수문자, 공백이 있어 잘못 입력되었으며 Age 변수의 4 번째 행의 cell 에 숫자가 아닌 문자인 “?” 가 입력되었다.



	A	B	C	D	E	F
1	ID No.	이름	성별	Age	Wt (kg)	
2	101	Park	2	25	65.3	
3	102	Lee	1	54	55.7	
4	103	Kim	1	?	48	
5	104	Oh	2	40	60.5	
6	105	Jung	1	30	51.2	
7						

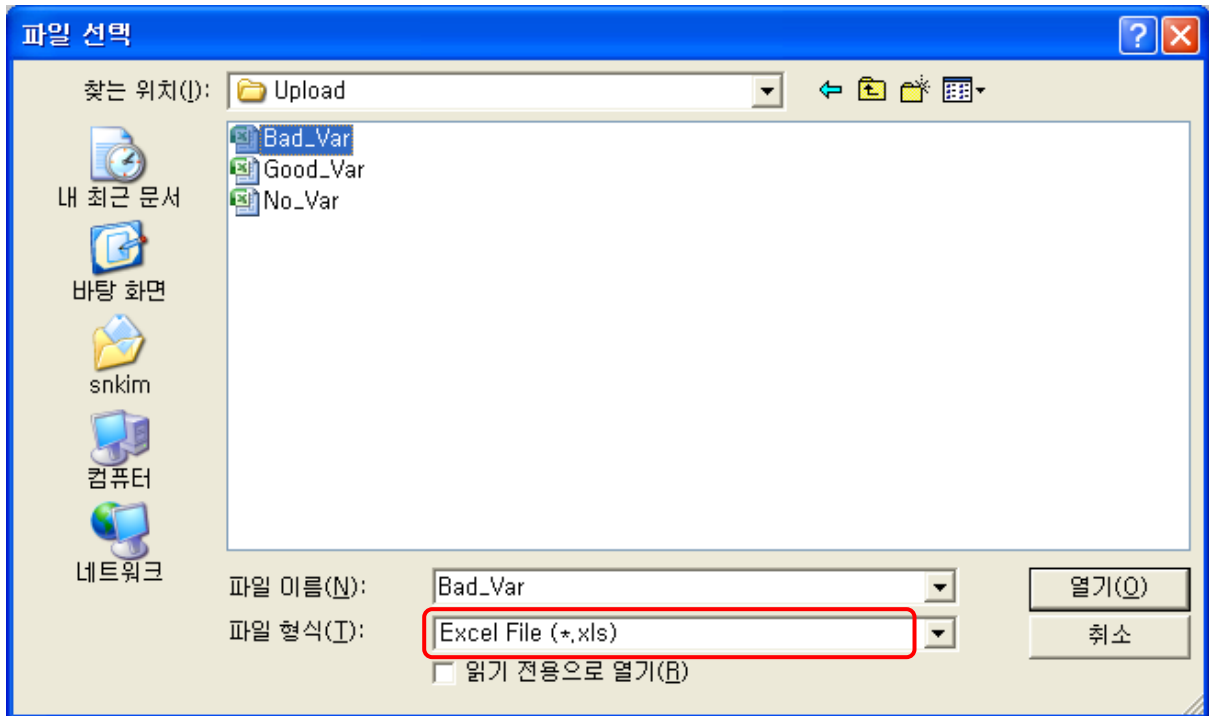
dBSTAT 에서 Excel 파일 열기

 시작화면에서 [데이터베이스 사용]을 선택하거나 메뉴에서 [파일] ➔ [파일열기]를 클릭한다.



파일선택 창에서 파일형식 ➔ Excel File (*.xls) 선택한 다음 엑셀파일 이름을 클릭하여 파일을 연다.

이번 예제에 해당하는 엑셀파일인 Bad_Var 를 선택한다.



다음과 같이 엑셀파일이 dBSTAT 파일(Bad_Var.DBF)로 변환되었다.

Database: D:\WUpload\WBad_Var.DBF

	IDNO_	V2	V3	AGE	WT_KG_	
1	101	Park	2	25	65.30	
2	102	Lee	1	54	55.70	
3	103	Kim	1		48.00	
4	104	Oh	2	40	60.50	
5	105	Jung	1	30	51.20	

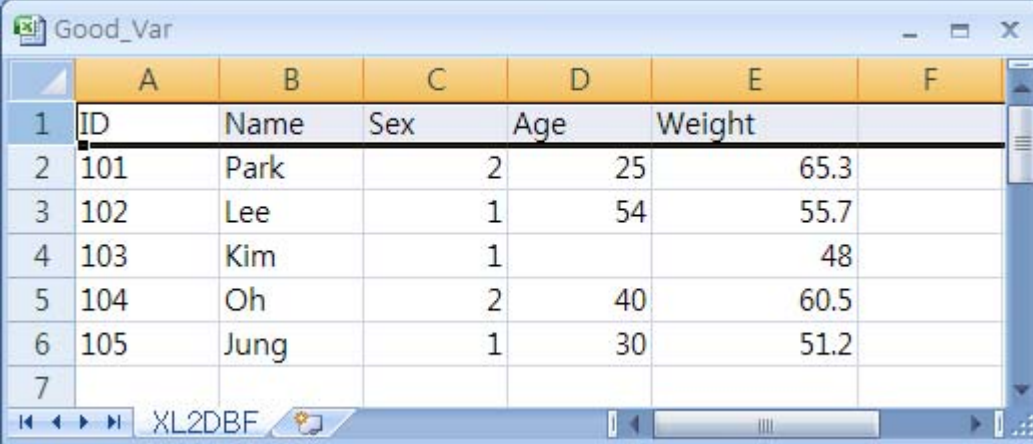
Database Table 1/5

엑셀에서 사용하였던 한글 변수명은 V2, V3 로 특수문자와 공백은 "_" 로 바뀌었음을 알 수 있다.

엑셀에서 입력되었던 "?" 문자는 빈칸(0)으로 보인다.

2. 엑셀 파일: Good Example

다음은 변수명과 입력이 올바른 예이다.



	A	B	C	D	E	F
1	ID	Name	Sex	Age	Weight	
2	101	Park	2	25	65.3	
3	102	Lee	1	54	55.7	
4	103	Kim	1		48	
5	104	Oh	2	40	60.5	
6	105	Jung	1	30	51.2	
7						

Good_Var.xlsx 파일을 dBSTAT 에서 열면 엑셀파일과 동일하게 보인다.

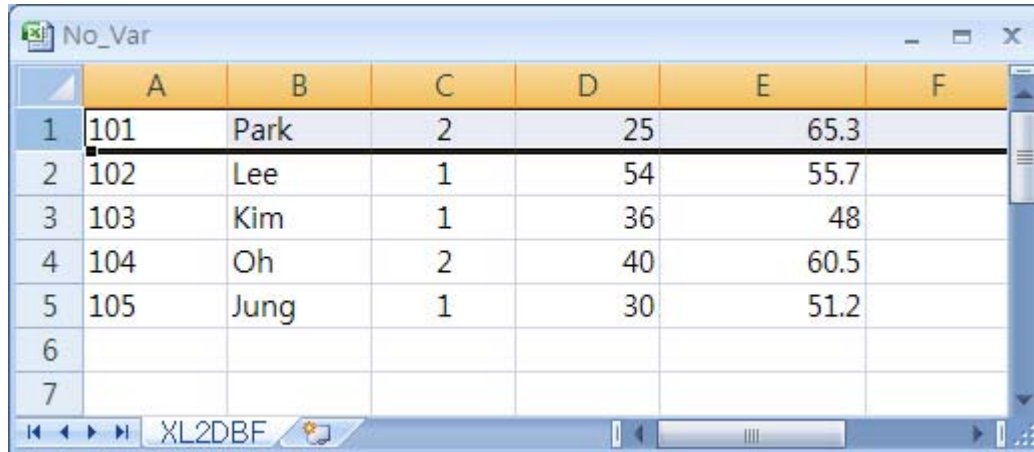


	ID	NAME	SEX	AGE	WEIGHT	
1	101	Park	2	25	65.30	
2	102	Lee	1	54	55.70	
3	103	Kim	1		48.00	
4	104	Oh	2	40	60.50	
5	105	Jung	1	30	51.20	

Age 변수에서 0 은 결측값으로 처리되어 빈칸으로 보이게 된다.

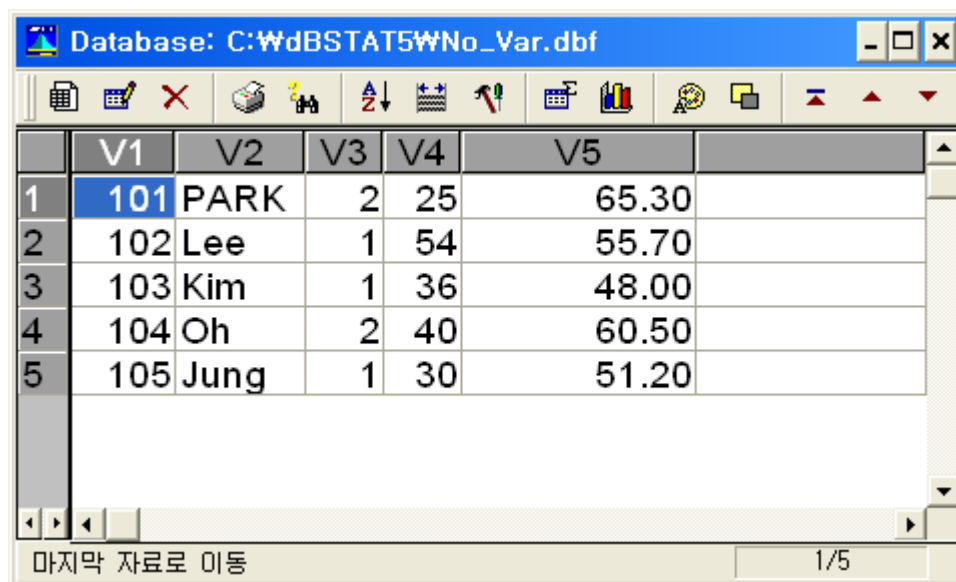
3. 엑셀 파일: No Variable

다음은 첫번째 행에 변수명이 없는 엑셀파일 예이다.



	A	B	C	D	E	F
1	101	Park	2	25	65.3	
2	102	Lee	1	54	55.7	
3	103	Kim	1	36	48	
4	104	Oh	2	40	60.5	
5	105	Jung	1	30	51.2	
6						
7						

No_Var.xlsx 파일을 dBSTAT 에서 열면 변수명이 V1, V2, ..., V5 로 생성되었다.



	V1	V2	V3	V4	V5	
1	101	PARK	2	25	65.30	
2	102	Lee	1	54	55.70	
3	103	Kim	1	36	48.00	
4	104	Oh	2	40	60.50	
5	105	Jung	1	30	51.20	

Excel 숫자 변수 → dBSTAT 문자 필드

엑셀에서 범주형인 숫자 변수(행)은 dBSTAT 에서 문자 필드로 바꾸어 주어야 한다.

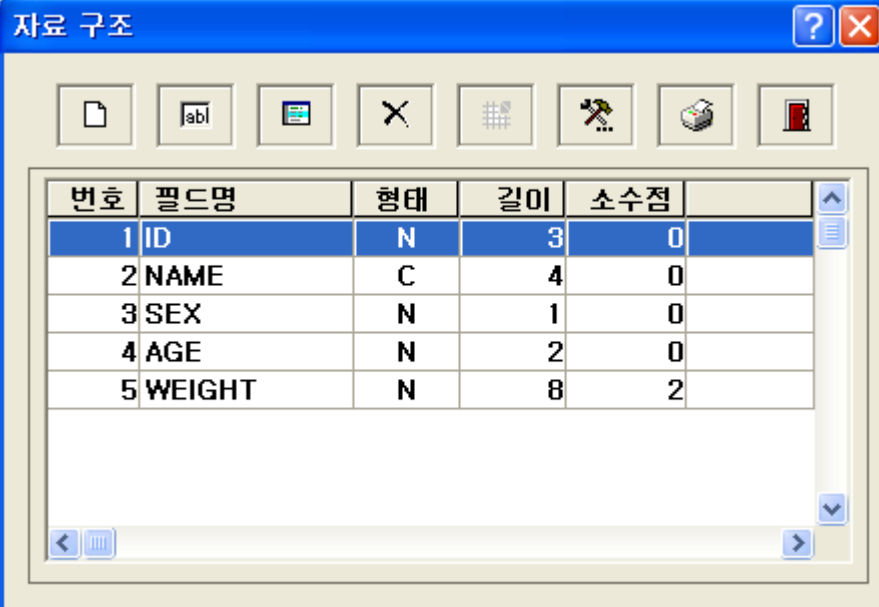


	ID	NAME	SEX	AGE	WEIGHT
1	101	Park	2	25	65.30
2	102	Lee	1	54	55.70
3	103	Kim	1		48.00
4	104	Oh	2	40	60.50
5	105	Jung	1	30	51.20

엑셀에서 SEX 변수가 숫자 변수이므로 dBSTAT 에서 통계분석에 사용하기 위해서는 문자 변수로 바꾸어야 한다.

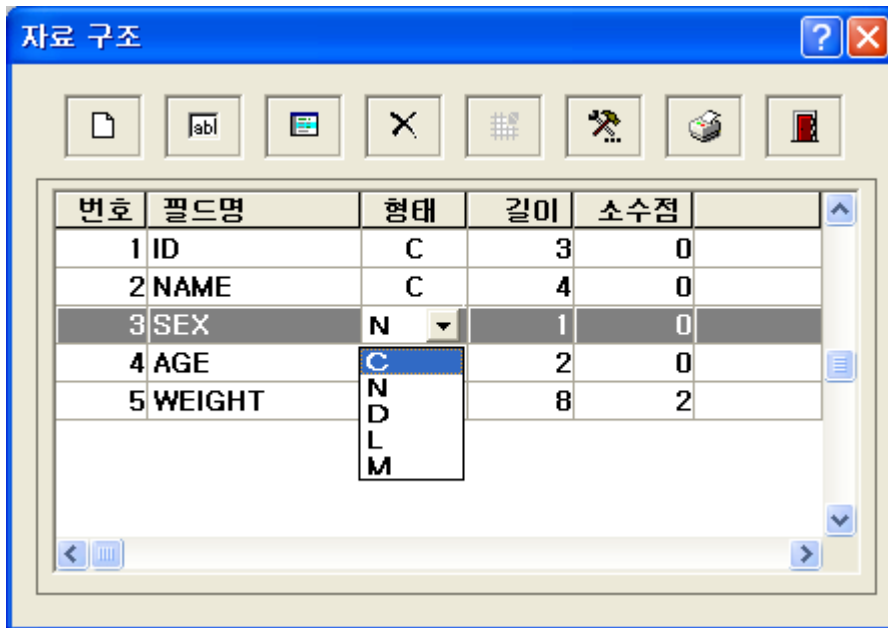


자료구조 버튼을 누르면 ID, SEX 필드가 숫자(N) 필드임을 보여준다.

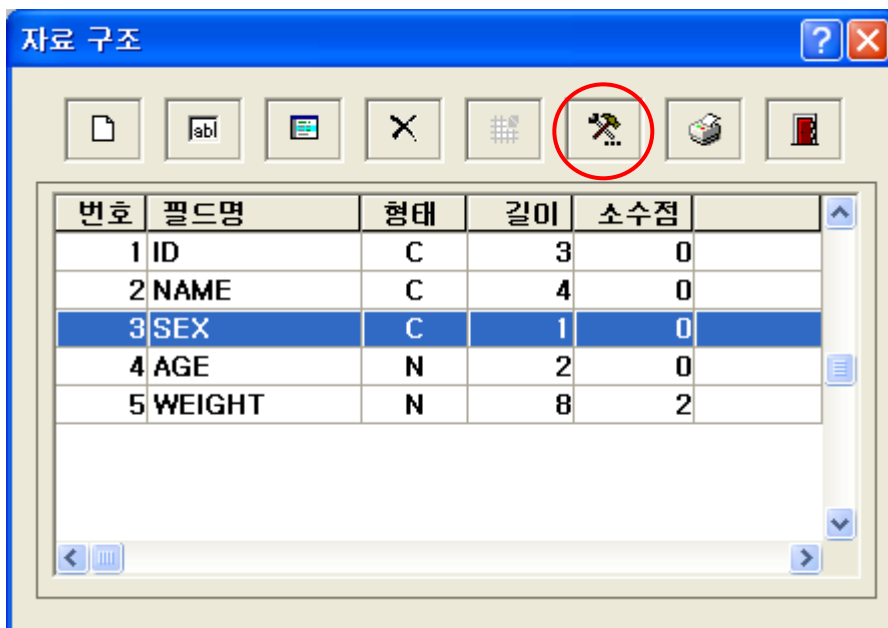


번호	필드명	형태	길이	소수점
1	ID	N	3	0
2	NAME	C	4	0
3	SEX	N	1	0
4	AGE	N	2	0
5	WEIGHT	N	8	2

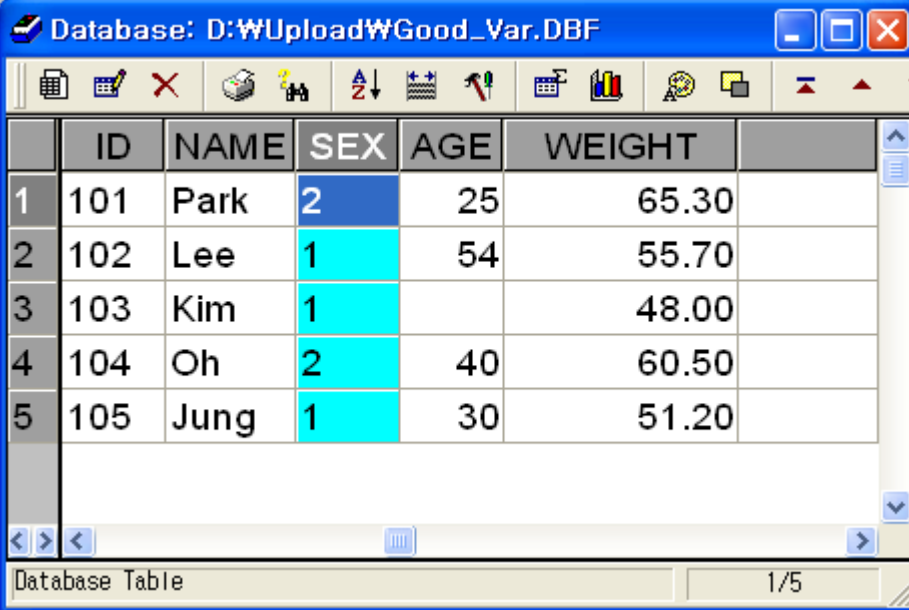
ID, SEX 필드형태를 C (문자) ➡ N (숫자)로 바꾼다.



[저장] 버튼을 클릭하여 자료구조 작업을 마친다.



문자 필드는 숫자가 좌측 정렬되어 왼쪽부터 보이게 된다.



	ID	NAME	SEX	AGE	WEIGHT	
1	101	Park	2	25	65.30	
2	102	Lee	1	54	55.70	
3	103	Kim	1		48.00	
4	104	Oh	2	40	60.50	
5	105	Jung	1	30	51.20	

ID,SEX 변수의 자료를 보면 문자 변수로 바뀌어 숫자가 좌측 정렬되었다.

dBSTAT 은 데이터베이스에서 변수 형태를 자동적으로 알아서 통계분석을 시행하게 되므로 반드시 SEX 와 같은 범주형 명목변수는 문자필드로 만들어야 한다. 엑셀에서 SEX 변수에 “M”, “F” 등 문자로 입력되어 있으면 자동적으로 문자 필드로 인식이 되지만 “1”, “2” 등 숫자로 입력이 되어 있으면 숫자 필드로 자동 변환되므로 dBSTAT 에서 자료구조를 문자필드로 바꾸어야 한다.