

## **Chapter 2. SAS 데이터셋 만들기**

# SAS 데이터 셋의 구조

## ❖ 변수(Variable)

- 데이터 셋의 한 열로써 구체적인 속성을 나타내는 자료값들의 집합
- 구성
  - 문자형 변수: 문자, 숫자, 특수문자 등 기호 등으로 이루어짐
  - 숫자형 변수: 숫자만으로 이루어짐

## ❖ 개체(Observation)

- 데이터 셋의 한 행
- 동일한 관찰 단위에 대한 모든 변수들의 자료값들의 집합

ID	NAME	GENDER	WEIGHT	HEIGHT
101	김철수	M	74	170
102	이영희	F	68	166
103	안수지	F	55	155
104	박민호	M	72	167
105		M	66	169

# 외부 텍스트 파일 읽어오기

## ❖ 명령어

- DATA 명령문: 원하는 데이터 셋의 이름 지정
- INFILE 명령문: 외부 텍스트 파일이 저장되어 있는 경로 및 파일이름 지정
- INPUT 명령문: 변수 이름과 열 번호를 지정하기 위해 사용
- LABEL 명령문: 변수에 대한 설명을 붙여두고자 할 때 사용(필수사항은 아님)

### 텍스트 자료의 예: 기업이미지.txt

```
011M111
022M333
034F331
044M332
054M111
065F2..
073M111
085F111
095M113
102F232
```

# 외부 텍스트 파일 읽어오기

코드북			
변수이름	내용	열 번호	코드값
Id	고객 번호	1-2	
Age	연령대	3	1=20대, 2=30대, 3=40대, 4=50대, 5=60대 이상
Gender	성별	4	M=남자, F=여자
Item1	노력	5	1=예, 2=보통, 3=아니오, .=모름/무응답
Item2	소비자 중요	6	
item3	신뢰성	7	

## 외부 텍스트 파일 읽기

```
DATA company;  
INFILE "c:\SAS_Programming\Raw_Data\기업이미지.txt";  
INPUT id 1-2 age 3 gender $ 4 item1 5 item2 6 item3 7;  
LABEL id=고객번호 age=나이 gender=성별 item1=노력 item2=소비자 item3=  
신뢰;  
RUN;
```

# SAS 데이터 셋의 이용과 저장

## 프로시저의 사용 예

```
PROC FREQ DATA=company;  
TABLE age gender item1 item2 item3;  
RUN;
```

## ❖ 라이브러리를 이용한 SAS 데이터 셋의 저장

- SAS는 특별한 지정이 없는 경우 SAS를 구동한 후에 생성된 데이터 셋들은 WORK 라이브러리에 임시로 보관하였다가 SAS가 종료되면 모두 삭제함
- 생성된 SAS 데이터 셋을 특별한 라이브러리에 저장해 두게 되면 언제든지 생성된 데이터 셋을 사용할 수 있음
- 먼저 데이터 셋이 저장되는 폴더를 SAS 라이브러리로 등록시켜야 함

## LIBNAME 명령문

```
LIBNAME library 'folder';
```

# SAS 데이터 셋의 이용과 저장

## LIBNAME 명령문의 사용 예

```
LIBNAME mysas 'c:\SAS_Programming\SAS_Data';  
DATA mysas.company;  
INFILE 'c:\SAS_Programming\Raw_Data\기업이미지.txt';  
INPUT id 1-2 age 3 gender $ 4 item1 5 item2 6 item3 7;  
LABEL id='고객번호' age='나이' gender='성별' item1='노력' item2='소비  
자' item3='신뢰';  
RUN;
```

- LIBNAME 명령문은 특정 폴더의 경로에 대해서 라이브러리라는 임시의 별칭을 붙여주는 것
- LIBNAME 명령문에 의해 생성된 라이브러리를 이용하여 SAS 데이터 셋을 물리적인 파일로 디스크에 저장함
  - 단, LIBNAME 명령문에 지정되는 폴더는 하드디스크상에 이미 존재하고 있어야 함
  - 폴더 안에 company.sas7bdat라는 파일이 생성되어 있음
- LIBNAME 명령문에 의해 등록된 라이브러리는 SAS가 구동되는 동안에만 사용되는 임시 별칭임
- 새로 SAS를 구동하는 경우에는 다시 LIBNAME 명령문을 수행해야 함

# SAS 데이터 셋의 이용과 저장

## ❖ 라이브러리 탐색기의 이용

- 라이브러리 탐색기를 이용하면 라이브러리를 메뉴 방식으로 관리할 수 있음
- 라이브러리를 계속 등록하여 사용할 수 있으므로 편리함
- SAS 데이터 셋의 내용 보거나 변수 이름의 변경 등 데이터 셋을 전반적으로 관리 할 수 있음
- 새로운 라이브러리의 생성
  - 라이브러리 생성 아이콘을 클릭-> 새로운 라이브러리 대화상자에서 라이브러리 이름과 폴더의 경로 지정
  - 시작할 때 자동 할당: 다음에 SAS를 구동할 때도 라이브러리의 지정이 계속 유지됨

# SAS 데이터 셋의 이용과 저장

- SAS 데이터 셋의 탐색: 폴더에 있는 데이터 셋을 탐색하고 편집 가능
  - 라이브러리 탐색기에서 해당 라이브러리 아이콘 더블클릭
  - 해당 데이터 셋을 더블클릭
  - View Table 윈도우
    - [보기] 메뉴: 데이터 셋 보기에 관한 여러 가지 옵션 변경
    - [데이터] 메뉴: 데이터 셋의 일부 내용만 보거나 정렬 및 검색 등 수행



# SAS 데이터 셋의 이용과 저장

## ❖ 라이브러리의 호출

- LIBNAME 명령문을 통해서 등록된 SAS 데이터 셋을 지칭할 경우
  - ‘라이브러리 이름.데이터 셋 이름’
- 라이브러리 이름 부분을 생략할 경우 임시 라이브러리인 WORK를 지정하는 것과 같음

### 라이브러리 호출: 프로시저 단계

```
PROC FREQ DATA=mysas.company;  
TABLE age gender item1 item2 item3;  
RUN;
```

### 라이브러리 호출: 데이터 단계

```
LIBNAME mysas "c:\SAS_Programming\SAS_Data";  
LIBNAME new "c:\temp\newdata";  
  
DATA new.company1;  
    SET mysas.company;  
    itemmean=(item1+item2+item3)/3;  
RUN;
```

# 데이터 가져오기 마법사 이용하기

# IMPORT 프로시저 이용하기

## ❖ IMPORT 프로시저

- 데이터 가져오기 마법사의 명령어
- PROC IMPORT 명령문
  - DBMS=옵션: 가져올 파일의 형식을 지정
  - DATAFILE=옵션: 파일이 저장되어 있는 경로와 이름을 지정
  - OUT=옵션: 데이터 셋이 저장될 라이브러리와 데이터 셋 이름을 지정
  - REPLACE 옵션: 라이브러리 내에 동일한 이름의 데이터 셋이 있는 경우 이를 덮어쓰도록 지정
  - RANGE=명령문: Excel파일 안에서 데이터를 가져올 워크시트를 지정하는 것
  - FMTLIB=명령문: SPSS파일에 변수값 레이블이 포함되어 있는 경우 이를 저장하여 사용하고자 할 때 지정

## IMPORT 프로시저

```
PROC IMPORT DBMS=옵션 DATAFILE=옵션 OUT=옵션 REPLACE;
```

# IMPORT 프로시저 이용하기

## IMPORT 프로시저: Excel 데이터 가져오기

```
PROC IMPORT DBMS=EXCEL  
DATAFILE='c:\SAS_Programming\Raw_Data\기업이미지.xls'  
OUT=mysas.company  
REPLACE;  
RANGE='Sheet1$';  
RUN;
```

## IMPORT 프로시저: SPSS 데이터 가져오기

```
PROC IMPORT DBMS=SPSS  
DATAFILE="c:\SAS_Programming\Raw_Data\기업이미지.sav"  
OUT=mysas.company  
REPLACE;  
FMTLIB=mysas.company_formats;  
RUN;
```

# 원자료를 프로그램 내부에 입력하기

- 원자료의 개체 수가 크지 않을 경우 SAS 프로그램의 데이터 단계에서 원자료를 직접 입력하여 데이터 셋을 생성하는 것이 효율적
- CARDS; 뒤에 위치하는 모든 자료값들은 ;(semicolon)이 나타날 때까지 원자료로 인식

## CARDS 명령문

```
CARDS;  
데이터 입력  
;
```

## 원자료를 프로그램 내부에 입력하기

```
LIBNAME mysas 'c:\SAS_Programming\SAS_Data';  
DATA mysas.htwt;  
INPUT name $ gender $ dept $ age height;  
LABEL name='이름' gender='성별' dept='학과' age='나이' height='키';  
CARDS;  
김철수 M STAT 25 170  
강민호 M STAT 20 169  
;  
RUN;
```

# 연습문제

❖ 아래의 자료를 텍스트 파일로 저장하고, 이 텍스트 파일을 INFILE 명령문을 이용하여 SAS 데이터 셋으로 저장하여라. 또한 'c:\w'에 example 폴더를 생성하고 이 폴더를 'ex'라는 이름의 SAS 라이브러리로 생성하여 위의 데이터를 저장하여라.

변수	내용	열 번호	텍스트 자료
Obs	고객 번호	1-2	1 Monona Steve 32 2 Milwaukee Tom 44 3 Madison Kim 25
City	도시	3-12	
Name	성명	13-18	
Age	나이	19-20	

❖ 아래의 데이터셋을 엑셀 파일로 저장하고 IMPORT 프로시저를 이용하여 저장한 엑셀 파일을 'ex' 라이브러리에 저장하여라.

Region	State	Month	Expenses	Revenue
Southern	GA	JAN2001	2000	8000
Southern	GA	FEB2001	1200	6000
Southern	FL	FEB2001	8500	11000
Northern	NY	FEB2001	3000	4000
Northern	NY	MAR2001	6000	5000
Southern	FL	MAR2001	9800	13500
Northern	MA	MAR2001	1500	1000