

## 2-3강

### 통계데이터

- 통계 분석용 데이터는 변수와 개체라는 두개의 차원에 의해 매트릭스 형태로 배열된 전체 자료의 집합

ID	SQ1	SQ2	SQ3	SQ4	Q01	Q02	Q03
A001	1	2	1	1	3	5	7
A002	1	3	1	2	4	5	6
A003	2	1	2	1	2	3	4
A004	2	2	3	1	4	3	2
A006	1	1	2	1	2	4	6
B001	3	2	3	3	2	5	6
B002	1	2	2	1	3	4	5

	name	clas	comp	stat
1	sohn	1	90	80
2	park	1	70	80
3	kim	2	50	40
4	lee	1	60	90
5	song	2	70	70
6	moon	2	80	70

Pg 51

### SAS Data Sets의 구조

OBS	ID	NAME	SEX	WEIGHT	HEIGHT
1	101	김철수	M	74	170
2	102	이영희	F	68	166
...	...	...	...	...	...
6	106	강지영	F	52	.

- 변수 (Variable)

데이터셋의 한 열로써 구체적인 속성을 나타내는 자료값들의 집합.

문자형 변수, 숫자형 변수와 날짜형 변수로 구분됨.

- 개체 (Observation)

데이터셋의 한 행으로써 동일한 관찰단위에 대한 모든 변수들의 자료값들의 집합.

### SAS Data Sets

- SAS데이터 셋은 SAS시스템에 의해 만들어진 매트릭스형 자료들의 집합
  - ※ 관계형 DBMS나 통계분석용 Data 테이블도 유사한 구조

데이터 셋 이름, 생성엔진,  
생성일시 : 2013-03-06 10:00 AM

관찰치(레코드)수 : 6 개, 변수(필드) 수 : 4개

변수속성정의 : **NAME** 문자형(10), **CLAS** 숫자형(4)  
**COMP** 숫자형(8), **STAT** 숫자형(8)

참조인덱스 : hose\_stat\_idx

참조카탈로그 : hose\_format

**Descriptor**

인덱스필드  
clas

**indexes**

시스템 사용  
변수 리스트

**indexes**

	name	clas	comp	stat
1	sohn	1	90	80
2	park	1	70	80
3	kim	2	50	40
4	lee	변수 Variable Column Field	60	90
5	song		70	70
6	moon		80	70

개체  
**Obsevation(관측치)**  
**Row, Record**

**DATA**

1: 1,2,4  
2: 3,5,6

Pg 36

Descriptor에 변수속성정의가 쓰여있다. 이것을 확인하자.

→ NAME 문자형(10)는 NAME이라는 변수는 문자형이고, 10byte를 차지한다는 의미이다.

## 변수 속성

- 필수속성
  - 이름(name)
  - 유형(type)
  - 길이(length)
- 추가속성
  - 출력형식(format)
  - 입력형식(informat)
  - 레이블(label)

## 필수 속성

변수가 생성되려면 반드시 지정되어야 하며, 변수 생성 시점은 DATA step의 상단부터 읽어서,

가장 먼저 발견되는 시점에서 생성됨

## 추가 속성

사용자 필요에 의해 추가로 할당하여 사용할 수 있음

### 변수 속성 : 이름(name)

	최대길이	공통 및 특이사항	한글사용 가능 여부
라이브러리명	8byte	<ul style="list-style-type: none"><li>▪ Underscore(_) 또는 영문자(A-Z)로 시작하여야 하며 나머지는 underscore, 영문자(A-Z), 숫자(0-9)로 구성</li><li>▪ 대소문자 구분하지 않음</li></ul>	불가능
Data Set명	32byte		불가능
변수명			<ul style="list-style-type: none"><li>▪ 시스템 옵션 변경 시 가능</li><li>▪ 한글 또는 특수문자(공백,대쉬 등)을 사용할 경우는 SAS 시스템 옵션 VALIDVARNAME 을 ‘ANY’값으로 설정해야 함</li><li>▪ Options validvarname=any;</li></ul>
출력형식명			<ul style="list-style-type: none"><li>▪ 가능</li><li>▪ 사용자 정의 출력형식 생성시 출력형식 명은 숫자나.(period)로 끝날 수 없음</li></ul>

### 변수 속성 : 유형 및 길이

유형	길이	결측값 (missing value)	
문자(\$)	최소 1byte - 최대 32,767byte	Blank로 표시	모든 값(문자, 숫자, 특수기호, 공백 등)을 포함할 수 있음
숫자	기본 8byte	.(period)로 표시	날짜, 시간, 날짜시간 값은 숫자형 변수로 저장함

### 날짜 및 시간 관련

유형	기준	저장값	상수(constant) 지정 방법
날짜	1960년 1월 1일을 0으로 저장	기준 날짜와 저장하고자 하는 날짜간의 일수 차이 예) 1960년 1월 3일은 2로 저장	'ddmmmyyy'D 예) 2007년 9월 17일은 '17sep2007'
시간	자정을 0으로 저장	기준 시간과 저장하고자 하는 시간 간의 초 차이	'hh:mm:ss'T 예)오전 10시는 '10:00:00'
날짜 시간	1960년 1월 1일 자정을 0으로 저장	기준 값과 저장하고자 하는 날짜 시간값 간의 초 차이	예)2007년 9월 17일 오전 10시 '17sep2007:10:00:00'DT

## 변수 속성 : 추가 속성들

속성	기준	예시
출력 형식	변수가 저장하고 있는 값을 어떻게 표시하라는 지시	예) 12345 → 12,345 comma6. 0 → 1960-01-01 yymmdd10.
입력 형식	변수에 값을 저장할 때 어떻게 저장하라는 지시	예) 01jan1960 → 0 Date9. 홍길동 → 홍길동 \$6.
레이블	출력물에서 변수 이름 대신에 사용되며, 256byte까지 지정될 수 있고, 특수문자와 한글을 사용할 수 있음	

yymmdd10.에서 10은 자릿수를 의미한다.

yymmdd8.을 하면 -가 다 빠진다.

yymmdd6.을 하면19600101에서 19가 빠져 600101이 된다,

ddmmmyy은 date를 의미하는 것이다.

\$6.은 6byte를 의미한다.

- SAS 변수 속성(요약)

속성	KEYWORD	용도 및 규칙
이름	VARNAME	영문자와 “_” 로 시작, 2번째 부터는 숫자도 가능, 32자
유형	VARTYPE	숫자형, 문자형(\$)
길이	LENGTH	문자형은 32,767자리 까지 가능 (default는 8) 숫자형은 8자리 고정 소수점이 기본 (4~32, default는 8)
인포맷	INFORMAT	숫자형(소수형, 정수형, 날짜형, 통화형), 문자형
포맷	FORMAT	숫자형(소수형, 정수형, 날짜형, 통화형), 문자형
레이블	LABEL	256문자, 변수설명
번호	VARNUM	INPUT 입력순서 > 할당문 순서 대로 번호 지정 (목록순)

## SAS 라이브러리

### LIBRARY 란

- SAS 파일 (dataset, catalog) 들이 저장되는 곳 또는
- 저장된 SAS 파일들의 복합체명칭 (Database명)
- 파일시스템에 존재하는 폴더 또는 디렉토리 (SAS에서 연결사용하는)

\* LIBRARY가 지정되어야만 SAS파일(데이터셋, 카탈로그)을 저장하거나 저장된 SAS파일을 참조하여 재사용할 수 있다.

\* 별도의 라이브러리가 지정되어 있지 않으면, 임시저장소에 저장해 두었다가 SAS 세션이 종료되면 모든 SAS파일을 삭제함

Aa Name	≡ 파일시스템	≡ 데이터베이스	≡ 스프레드시트	≡ SAS
<u>저장소</u> <u>명</u>	폴더/디렉토리	DB	0000.xls	LIBrary
<u>데이터</u> <u>파일</u>	파일	TABLE	SHEET	SAS파일 - 데이터셋(.sas7bdat) - 카탈로그(.sas7bcatalog)

SAS에서는 파일 시스템의 폴더명이나 디렉토리명을 직접 사용하지 않는다.

→ SAS만의 연결방법을 사용하여 지정된 폴더나 디렉토리를 이용함

- 라이브러리 정의

SAS파일이 저장된 디렉토리 또는 폴더

SAS 파일의 저장장소

참고) SAS 파일이란?

- SAS에 의해 생성 및 관리되는 파일
- 파일의 확장자가 sas7bxxx인 파일

예) SAS 데이터 셋 : xxx.sas7bdat

SAS 카탈로그 : xxx.sas7bcatalog

- SAS는 기본적으로 데이터셋, 포맷, 응용프로그램, 매크로 등을 정의할 때 기본적으로 WORK 라는 임시 폴더에 저장

(이 임시 폴더의 물리적 경로는 SAS 가동시 임의로 결정)

- 라이브러리 종류

- 시스템정의/사용자정의
  - 시스템정의 : SAS 설치 시 자동적으로 생성
  - SASHELP, SASUSER, WORK 등
- 사용자정의 : SAS 사용자가 생성

- 임시/영구

- 임시 : SAS 종료 시 자료가 자동 삭제
- 영구 : SAS 종료 후에도 라이브러리 내의 자료가 보존

WORK 외의 라이브러리

☞ SAS 데이터 셋은 [WORK .데이터 셋]과 같이 “WORK”라는 라이브러리가 생략되어 있음

☞ WORK 임시폴더는 휘발성이므로 저장되어 있던 모든 정보는 SAS의 종료와 함께 모두 삭제됨

☞ 사용자가 생성하고 정의한 데이터 셋을 영구적으로 저장하려면 특정 폴더에 저장할 수 있도록 라이브러리를 등록해야 함

## 라이브러리의 실제 사용

- SAS 프로그램의 일반형식

LIBNAME library( 라이브러리 이름 ) 'folder'( 경로 ) ;

- example

```
LIBNAME stat1 'F:\data';
```

- 생성된 라이브러리 경로의 확인은 다음과 같이 [LIST] 옵션을 사용

- example

```
LIBNAME stat1 list;  
run;
```

- txt파일을 infile하여 라이브러리에 저장하는 example 코드

```
DATA ex22;  
INFILE "/folders/myfolders/sasuser.v94/data/2_3/ex2.txt";  
input name $ num gender $ dept $ mid final;  
RUN;  
  
PROC PRINT DATA=ex22;  
RUN;  
  
data mylib.ex22;  
set ex22; /*라이브러리 mylib에 data ex22복사하기*/  
run;  
  
proc print data=mylib.ex22;  
run;
```

