

# 3-1강

## SAS프로그래밍의 구조

### 1. DATA STEP

- SAS 데이터셋에 데이터 입력하기
- 새로운 변수값 계산
- 데이터의 오류 확인 및 수정 (외부데이터가 이상할 수도 있기에 확인을 해야한다.)
- 기존 데이터셋의 서브셋, 병합, 업데이트 등으로 새로운 SAS 데이터셋 만들기

### 2. PROC STEP

- 리포트 출력
- 기술 통계 생성
- 테이블 리포트 작성
- 도표 및 차트 생성

## Data 단계(Data Step)

### 1. 시작

DATA <SAS data set names> <options>

### 2. data step에는 일반적으로 다음 중 하나의 문장이 있다

INPUT(내가 변수를 정해준다), SET(data를 복사), MERGE(data 결합),  
UPDATE(data 결합)

### 3. 기본적인 data step 문장

```
DATA one;                                /*data step의 시작*/
INPUT id gender $ weight;                /*자료를 읽음*/
CARDS;                                  /*자료의 시작*/
1234 M 49.5
1537 F 40.0
1745 M 70.2
1955 F 42.3                             /*변수가 3개, observation이 4개*/
```

```
;                                /*자료의 끝*/
RUN;
```

VIEWTABLE: Work.One			
	id	gender	weight
1	1234	M	49.5
2	1537	F	40
3	1745	M	70.2
4	1955	F	42.3

→ 이 data는 work라는 임시 라이브러리에 저장된다.

## Data 입력 방법 (1. 자유 입력\_List Input)

data가 공백으로 구분되어 있으면 사용한다.

- Input문과 Cards(또는 datalines)문 사용
    - Input - \$는 그 관측값이 문자일 때
    - Cards - 빈칸으로 관측값을 구분한다
- Input 자료의 개수만큼만 입력 받는다.

```
DATA one;                                /*DATA  data셋 명;*/
INPUT id gender $ weight;                /*INPUT  변수의 이름;*/
CARDS;
1234 M 49.5
1537 F 40.0
1745 M 70.2
1955 F 42.3
;
RUN;
```

- 자료의 출력

```
PROC PRINT DATA=one;
RUN;
```

- RUN 문 : 각 DATA 단계 또는 PROC 절차의 입력이 완료되었음을 SAS 시스템에 알리는 역할

## Data 입력 예 (중요하다 이 말이야)

```
data one;
  input x y z;
  cards;
  1 2 3
  4 5 6
  ;
run;

data two;
  input x y z;
  cards;
  1 2 3 4 5 6
  7 8 9 8 4 2
  ;
run;

data three;
  input x y z;
  cards;
  1 2
  3 4 5
  6 7 8
  ;
run;

data four;
  input x y z;
  cards;
  1 2 3 4
  5 6
  7 8
  9 0 1
  ;
run;
```

OBS	x	y	z
1	1	2	3
2	4	5	6

OBS	x	y	z
1	1	2	3
2	7	8	9

OBS	x	y	z
1	1	2	3
2	6	7	8

OBS	x	y	z
1	1	2	3
2	5	6	7
3	9	0	1

## Data 입력 방법 (2.열 지정 입력\_Column Input)

- Input문과 Cards(또는 datalines)문 사용
- Input - \$는 그 관측값이 문자일 때
  - Column input은 자료가 빈칸으로 구분되어 있지 않거나 몇 개의 자료를 건너 띄고 필요한 자료만을 읽을 때 쓰인다.

```
DATA one;                                /*data 데이터이름*/
INPUT id 1-4 gender $ 5-6 weight 7-12; /*Input 자료의 이름*/ /*숫자 → 데이터 입력 공간*/
CARDS;
1234 M 49.5
1537 F 40.0
1745 M 70.2
1955 F 42.3
;
RUN;
```

숫자 → 데이터 입력 공간

1	2	3	4	5	6	7	8	9	0	1	2
1	2	3	4		M		4	9	.	5	
1	5	3	7		F		4	0			
1	7	4	5		M		7	0	.	2	
1	9	5	5		F		4	2	.	3	

## Data 입력 방법 (3.포맷 입력\_Formatted Input)

→ 중간고사 시험에도 자주 나온다.

- format : a.b ; a는 전체 자릿수, b는 소수점 밑자리수  
주어진 자리만큼 순서대로 읽어간다.  
a개의 자릿수만을 읽고, 그 다음에 b자리수 만큼 소수점을 찍겠다는 의미  
ex) 4.2이면 12345 → 12.34가 된다.
- a. = a.0 → 정수 자릿수만 읽겠다는 의미
- 자료에서 소수점이 주어진다면 format에서의 소수점 밑자리수는 무의미해진다.  
ex) 4.2인데, 주어진 data가 3.12이면 그냥 3.12 그대로 간다.

```
data three;  
  input id 4 gender $ 3. weight 4.1;  
  card;  
1234 M 49.5  
1537 F 40.0  
1745 M 70.2  
1955 F 42.3  
;  
run;
```

1	2	3	4	5	6	7	8	9	0	1	2
1	2	3	4		M		4	9	.	5	
1	5	3	7		F		4	0			
1	7	4	5		M		7	0	.	2	
1	9	5	5		F		4	2	.	3	

	id	gender	weight
1	1234	M	49,5
2	1537	F	40
3	1745	M	70,2
4	1955	F	42,3

```
data three;
  input x 4.2 a$ 4.;
  card;
1234A
12.3 A
  77 A B
199 A B
  199A B
;
run;
```

1	2	3	4	5	6	7	8
1	2	3	4	A			
1	2	.	3		A		
	7	7		A		B	
1	9	9			A		B
	1	9	9	A			B

VIEWTABLE: Work.Three		
	x	a
1	12,34	A
2	12,3	A
3	0,77	A B
4	1,99	A B
5	1,99	A B

## 예5.8) 하나의 개체가 여러 줄로 입력된 자료

- 하나의 관찰개체의 자료가 여러 줄에 걸쳐서 입력될 때 사용
  - / : 포인터의 위치를 다음 줄의 첫 열로
  - #n : 줄 포인터 - 포인터의 위치를 n번째 줄의 첫 열로
  - @n : 열 포인터 - n번째 열로 자료의 입력시점을 이동

```
DATA club;
INPUT indo name $ 6-19
  / team $6.          /* /는 다음줄 첫번째부터 읽으라는 의미이다. */
  #3 strtwght endwght;
```

```
CARDS;
1023 David Show
Red
189 165
1049 Amelia Serrano
Yellow
189 165
;

RUN;
```

VIEWTABLE: Work.Club					
	indo	name	team	strtwght	endwght
1	1023	David Show	Red	189	165
2	1049	Amelia Serrano	Yellow	189	165

## Data 입력 방법: Infile 문

: 외부 파일로 부터 데이터셋의 생성

- 자료가 외부 파일에 저장되어 있을 경우 Cards문을 사용하여 자료를 입력하지 않고 직접 불러 들일 때 사용

Input – 자료들의 이름(변수)을 지정

```
data test;
infile "/folders/myfolders/sasuser.v94/data/2_3/ex2.txt";
input x y z; /*column확인을 해줘야 함*/
run;
```

	x	y	z
1	1	2	3
2	4	5	6
3	7	8	9

## Data 읽기

- 기존 SAS Data Set 읽기(편집기 사용)
  - DATA 문장
  - Data Step 시작 문장

- DATA 키워드 옆에 생성될 SAS Data Set 이름을 적음
- SET 문장 - 새로 결합, copy의 의미를 담고 있음
  - 기존 SAS Data Set을 그대로 읽을 때 사용하는 문장
  - SAS Data Set 외의 Raw data 파일이나 기타 데이터 파일은 읽어올 수 없음
  - 기본으로 입력 SAS Data Set의 모든 변수와 모든 관측치를 읽어옴

```
LIBNAME 라이브러리 이름 '경로';
```

```
DATA 출력-SAS-data-set;
  SET 입력-SAS-data-set;
RUN;
```

```
LIBNAME korea "c:\sas"; /*[korea 라이브러리에 있는 sales 데이터를 불러들여 */
DATA work.subset1;      /*work라이브러리에 있는 subset1이라는 데이터명으로 */
  SET korea.sales;      /*저장]*/
RUN;
```

```
Data class;
INPUT name $ num gender $ dept $ mid final; # 변수 6개
CARDS;
Minyoung 1 F stat 50 50
Minji 2 F computer 50 50
...
;
RUN;
```

```
DATA new;      # 변수 7개
SET class;
score=mid+final;      # score변수 추가
RUN;
```

```
DATA new1;    # 변수 8개
SET new;
average=score/2;      # average변수 추가
RUN;
```