

5-2강

DATA Step에서 관측치 선택

- WHERE where-표현식;
- IF 조건표현식;
- IF 조건표현식 THEN DELETE

WHERE 문장

```
WHERE where-표현식;
```

- 입력 Data Set에 대하여 읽어 올 관측치에 대한 조건을 지정
- PROC Step에서도 사용 가능 (IF는 PROC step에서 안 된다.)
- Where-표현식은 연산자와 피연산자의 조합으로 원하는 조건을 기술
 - 피연산자 : 변수 또는 상수 사용 가능.
 - 단, 변수를 사용하는 경우 반드시 입력 Data Set에 존재하는 변수이어야 함
 - 연산자 : 비교/논리/특수 연산자 사용 가능
- 입력 Data Set에 대하여 읽어 올 관측치에 대한 조건을 지정

```
DATA work.subset1;  
  SET orion.sales;  
  where country="AU" and  
    Job_Title contains "Rep"; /*contains -> Rep를 포함하는 Job_Title이다.*/  
RUN;
```

산술연산자

- SAS 산술연산자는 자료의 산술계산을 행하는데 필요
- ex)
 - where (Salary / 12) < 6000;

- where (Salary / 12) * 1.10 >= 7500;
- where (Salary + Bonus) <= 10000;

Comparison and Logical Operator

Symbol	Mnemonic Operator	Meaning
=	EQ	equal to
~=, ^=	NE	not equal to
>	GT	greater than
<	LT	less than
>=	GE	greater than or equal to
<=	LE	less than or equal to
&	AND	and
, !	OR	or
	IN	contain

비교연산자

- ex)
 - where Gender = 'M';
 - where Gender eq ' ';
 - where Salary ne . ;
 - where Salary >= 50000;
 - where Country in ('AU','US') ; /*값들은 콤마나 공백으로 구분*/
 - where Country in ('AU' 'US') ; /*값들은 콤마나 공백으로 구분*/

WHERE 특수연산자

기호	연상기호	뜻
	BETWEEN (a) AND (b)	(a)와 (b) 사이의 자료 선택, 경계선 (a)와 (b) 포함
	IS NULL	결측값 선택
	IS MISSING	결측값 선택
	CONTAINS	해당 문자열 포함할 경우 선택
	LIKE (%)	해당 문자 패턴일 경우 선택

- ex)
 - where Job_Title contains 'Rep';
 - where Name like '%N'; /*끝에 N만 있으면 됨. %에는 뭐든지 와도 됨 */
 - where Name like 'T_M%'; /*_에는 한글자이면 뭐든지 와도 된다. */
- 활용)

Diana, Diane, Dianna, Dianthus, Dyan

Aa 유형	선택
<u>like 'D_an'</u>	Dyan
<u>like 'D_an_'</u>	Diana, Diane
<u>like 'D_an_'</u>	Dianna
<u>like 'D_an_'</u>	Diana, Diane, Dianna, Dianthus, Dyan

Smith , Smooth , Smothers , Smart , Smuggle

Aa 유형	선택
<u>like 'Sm%'</u>	Smith , Smooth , Smothers , Smart , Smuggle .
<u>like '%th'</u>	Smith , Smooth .
<u>like 'S__gg%'</u>	Smuggle .
<u>like 'S_o'</u>	a three-letter word, so it has no matches here.
<u>like 'S_o%'</u>	Smooth , Smothers .
<u>like 'S%th'</u>	Smith , Smooth .

부분집합화(Subsetting) IF 문장

```
IF 조건표현식 ;
```

- 기존의 Data Set 또는 현재 입력되는 관측값 중에서 (조건표현식)이 참이 되는 관측값만 선별하여 Data Set을 만들
- Data Set에서만 사용 가능
- Where 특수 연산자는 사용불가
 - 변수값(gender) 이 여자(f)인 경우만 실행
IF gender= "f" ;
 - 변수값(age) 이 missing 이 아니거나 0 이 아닌 경우만 실행
IF age

DELETE 문장 : 부분집합화 IF 문의 반대개념

```
IF 조건표현식 THEN DELETE;
```

- 보통 조건 IF와 함께 사용되며, 조건표현식이참일 경우 해당 관측치는 삭제됨
- DELETE 문장을 중복해서 사용하는 경우는 모든 관찰치가 삭제될 수 있으니 주의해야 한다.
- DELETE 문의 실행결과가 부분집합화 IF 문의 실행결과와 반대로 나타나기 때문에 SAS 프로그래밍에서는 부분집합화 IF 문을 많이 사용한다.

관측치 선택방법 비교

```
DATA work.newsal;  
SET orion.sales;  
WHERE Salary > 30000;  
NewSalary = Salary * 1.1;  
RUN;
```

```
DATA work.newsal;  
SET orion.sales;  
NewSalary = Salary * 1.1;  
IF NewSalary > 33000;  
RUN;
```

```
DATA work.newsall;
SET orion.sales;
NewSalary = Salary * 1.1;
IF NewSalary <= 33000 THEN DELETE;
RUN;
```

→ 똑같은 결과를 가져온다.

KEEP(또는 DROP) 문장

- 출력 Data Set에서 선택(제거)하고자 하는 변수를 지정

```
KEEP 선택할 variable - list;
```

```
DROP 제거할 variable - list;
```

```
Data work.subset1;
SET orion.sales;
  WHERE Country="AU" and Job_Title contains "Rep";
  KEEP First_Name Last_Name Salary Job_Title Hire_Date;
RUN;
```

```
DATA a;
INPUT name $ score1-score20;
avg = MEAN(OF score1-score20);
total = SUM(OF score1-score20);
KEEP name avg total;
CARDS;
...
RUN;
```

```
DATA a (KEEP name avg total);
INPUT name $ score1-score20;
avg = MEAN(OF score1-score20);
total = SUM(OF score1-score20);
CARDS;
...
RUN;
```

LABEL 문장

```
LABEL 변수 = "라벨";  
    변수 = "라벨" ... ;
```

- 변수의 라벨 속성을 부여
- 라벨은 최대 256자까지 가능

```
DATA work.subset1;  
set orion.sales;  
WHERE Country="AU" and Job_Title contains "Rep";  
KEEP First_Name Last_Name Salary Job_Title Hire_Date;  
LABEL Job_Title="Sales Title" Hire_Date="Date Hired";  
RUN;
```

FORMAT 문장

```
FORMAT 변수명 [형식] ;
```

- 변수의 출력형식 속성을 부여
- 출력형식은 변수값을 어떻게 보여주라는 지시임
- 실제 변수값에는 변경이 없음

```
DATA work.subset1;  
set orion.sales;  
WHERE Country="AU" and Job_Title contains "Rep";  
KEEP First_Name Last_Name Salary Job_Title Hire_Date;  
LABEL Job_Title="Sales Title" Hire_Date="Date Hired";  
FORMAT Salary comma8. Hire_Date yymmdd10.;  
RUN;
```

출력형식

```
<$> format <w>.<d>
```

\$	문자형 출력형식 이 기호가 있는 경우 문자형 변수에만 사용 가능
<i>format</i>	출력형식 이름
<i>w</i>	보여줄 총 길이(특수문자를 포함한 길이)
.	반드시 써야 함
<i>d</i>	숫자형 출력형식에서 소수점 아래 자리수

출력형식 예

Format	Stored Value	Displayed Value
\$4.	Programming	Prog
12.	27134.2864	27134
12.2	27134.2864	27134.29
COMMA12.2	27134.2864	27,134.29
DOLLAR12.2	27134.2864	\$27,134.29

기타 변수 속성 할당 문장

RENAME

- 원래 변수명 = 바꾸고자 하는 새로운 변수명
- 데이터 셋 a를 그대로 읽어 total 변수이름을 sum으로, avg 변수이름을 mean으로 변경

```
DATA b;
  set a;
  RENAME total=sum avg=mean;
RUN;
```

TITLE

- SAS 결과를 프린트할 때 프린트 윗부분에 제목을 쓰는 기능으로 10 줄까지 지정

```
TITLE[n] "title";
```

```
TITLE; /*이러면 타이틀 제목이 사라진다.*/
```

FOOTNOTE

- SAS 결과를 인쇄하고자 할 때 프린트용지 제일 아래 부분에 인쇄하는 기능으로 10줄 까지 지정
- 프린트용지의 제일 아래에서 몇 줄 위에 내용을 프린트 할 것인지를 지정(n : 해당줄)

```
FOOTNOTE[n] "text";
```

해지하고 싶으면

```
FOOTNOTE[n];
```

RETAIN

- 바로 직전의 DATA 단계에서 수행된 값을 유지시켜주는 역할
- 변수의 초기값을 지정할 때 주로 사용

```
DATA a1;  
INPUT x @@;  
CARDS;  
1 2 3 4 5  
;  
RUN;
```

```
DATA a2;  
SET a1;  
RETAIN sumx 0; /*초기치를 0으로*/  
sumx = sumx + x; /*매번 x누적치를 구함*/  
PUT sumx; /*로그창에 프린트*/  
RUN;
```

초기치 sumx 를 0 으로 주고 매번 x의 값에 대한 누적치를 구한다

retain문

- data step의 처음에 문장을 수행하기 전에 buffer의 내용을 결측값으로 만든다. 이 때 지정된 변수에 대한 buffer의 내용을 결측값으로 만들지 않으려면 retain을 사용하면 된다.

```
RETAIN variable-name [initial value];
```

```
DATA cusum;  
RETAIN cusum 0;
```

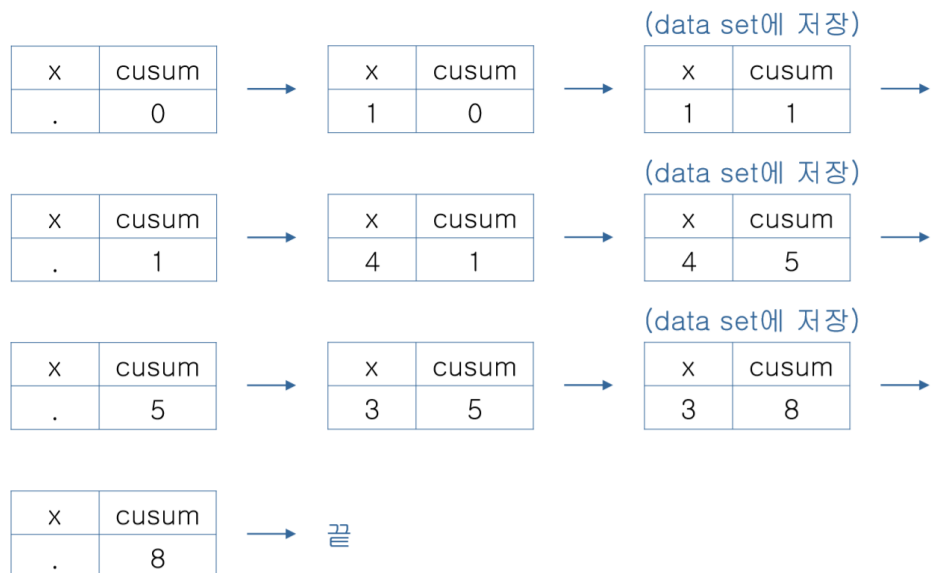


```
input x;
cusum = cusum + x;
CARDS;
1
4
3
;
```

결과:

cusum	x
1	1
5	4
8	3

buffer의 변화



MISSING

- 입력데이터 중 특수문자를 Missing 데이터로 간주하고자 하는 경우에 사용

```
MISSING values;
```

OPTION

- 시스템이 가지고 있는 기본값들을 잠정적으로 변경하고자 할 때 사용
- 보다 자세한 사항은 명령라인에서 OPTIONS라고 입력 후 엔터를 누르면 대화상자가 출력되고 그 곳에 여러 설정을 변경할 수 있음
- DATA에 앞서서 사용하는 기능임

선택사항에 사용할 수 있는 기능

- FIRSTOBS : 데이터 셋을 만들 때 몇 번째 자료부터 관측치 이용할 것인가를 지정
- OBS : 데이터 셋에 몇 개의 관측치를 이용할 것인가를 지정
- DATE SAS : 결과의 상단에 날짜를 프린트하고자 할 때 사용
- NODATE : SAS 결과의 상단에 날짜를 생략하고자 할 경우
- LINESIZE= : SAS 결과를 구할 때 결과의 폭을 지정
- MISSING= : 결측값을 다른 임의의 문자로 치환할 경우 사용
(예를 들어, MISSING=9는 모든 결측값을 9로 치환)
- PAGESIZE= : SAS 결과를 구할 때 한 페이지에 몇 행을 프린트할 것인가를 지정
- CENTER : 결과를 가운데 정렬
- NOCENTER : 결과를 왼쪽 정렬
- FMTSEARCH= : 포맷이 있는 위치를 지정 [예) FMTSEARCH=(WORK)]