



실험계획과 분석

심송용(한림대학교 데이터과학스쿨)

<http://jupiter.hallym.ac.kr>

SAS 기초

7 1 25.0 4 100.0

Frequency Missing = 2

이다.

@@

한 줄에 여러 케이스(case)의 자료가 있는 경우 INPUT 명령에서 @@를 붙인다.

줄맞춘 자료

만일 자료가 빈 칸이 없이 특정한 열에 줄을 맞추어 코딩되어 있으면

INPUT var format;

으로 사용하여 자료가 열을 맞춘 상태라는 것을 알려줄 수 있다. 이 경우는 인접하는 값들 사이에 빈 공간이 없어도 된다. 예를 들어

SAS 기초

```
INPUT GENDER $ 7 X2 8-10 Y 1-3;
```

라고 하면 변수

GENDER는 한 문자인 문자열 자료이며 일곱 번째 열에서 자료를 읽고

- X2는 8,9,10 열에 맞추어진 자료이며
- Y는 1,2,3 열에 맞추어진 자료인데 X2와 마찬가지로 숫자이다.

이 경우에 빈 칸은 숫자로 읽는 변수인 경우 0이 있는 것으로 생각한다. 만일 숫자로 읽을 변수 값이 소숫점이 필요하면 소숫점 이하의 자리수를 정할 수 있다. 예를 들어

```
INPUT GENDER $ 7 X2 8-10 1 Y 1-3 2
```

과 같이 소숫점 이하 자리수를 1과 2로 주면

- GENDER는 소수점이 없으며(이미 문자)
- X2는 8열에서 10열까지의 숫자를 읽어서 소숫점 앞에 8열과 9열에서 읽은 두 숫자를, 소숫점 뒤에 10열에서 읽은 숫자를 붙여서 숫자를 만든다. 이 때 원래 자료에 소숫점이 있

SAS 기초

으면 원래 자료의 소숫점이 우선한다.

- Y는 1-3열에 자료가 입력되었으며 1열은 소수점 왼쪽의 값, 2-3열은 소수점 오른쪽 값이다.

SAS 파일 안에 포함된 자료

SAS 명령문 안에 자료가 있을 때에는 CARDS; 명령과 ; 명령 사이에 자료를 입력한다.

자료 파일로 따로 만든 자료

만일 자료 파일이 따로 준비되어 있다면 이 자료 파일은 INFILE 문으로 읽을 수 있다.
INFILE 문은

```
INFILE file_name;
```

로 사용한다. 예를 들어

SAS 기초

INFILE 'a:/data/data1.dat';

와 같이 사용한다.

참고: INFILE 명령은 INPUT 명령보다 앞에 나와야 한다.

SAS Univserity Edition 내의 경로표시에 대한 주의

SAS Univserity Edition은 가상머신 내에서 실행되므로 가상 머신이 컴퓨터 내의 모든 리소스에 대한 접근을 하는 경우 보안상 위험에 노출되므로 사전에 허락된 곳만 접근할 수 있다.

이 위치는 최초 SAS University Edition 설치시 설정된 공유 폴더

SASUniversityEdition\myfolders에 국한된다. 즉,

- SAS University Edition에서 파일에 대한 접근은 폴더 SASUniversityEdition\folders 내에만 가능하므로 이 폴더 아래에 읽을 파일을 가져다 두어야 한다.
- SAS Univ. Ed. 의 경로 구분은 Unix 계열에서 사용하는 /이다.
- 한편 SASUniversityEdition 폴더는 SAS University Edition에서는 /folders로 인식되므

SAS 기초

로 SASUniversityEdition\myfolders\mydata\data.txt 라는 파일은

- /folders/myfolders/mydata/data.txt 로 설정한다.

산술연산과 내장함수

SAS에서의 연산은 대개의 컴퓨터 언어에서 사용하는 것과 비슷한 방법으로 한다.

사칙 연산은 +, -, * 및 /로 덧셈, 뺄셈, 곱셈 및 나눗셈을 한다. ** 는 지수이다. 즉

$Y = X^{**}3;$

의 결과로 만들어진 Y는 X의 세제곱이다.

SAS 기초

SAS 의 내장 함수

함수	함수 사용 결과
ABS(X)	X의 절대값
MAX(X)	X의 최대값
MIN(X)	X의 최소값
MOD(X, Y)	X를 Y로 나눈 나머지 값.
SIGN(X)	X의 부호. X의 값이 0이면 0.
SQRT(X)	X의 양의 제곱근. 음수인 X 값에 대해서는 .을 줌.
LOG(X)	X의 자연 로그
LOG2(X)	밑이 2인 X의 로그
LOG10(X)	밑이 10인 X의 로그
GAMMA(X)	Gamma 함수. 즉 $\int_0^\infty t^{x-1} e^{-t} dt$
DIGAMMA(X)	Digamma 함수. 즉 $\log'(\text{GAMMA}(X))$
EXP(X)	e^x .
NORMAL(<i>value</i>)	표준정규분포로부터의 난수. <i>value</i> 에는 0이나 5, 6, 7 자리 수의 홀수를 사용할 수 있다. (초기치로 사용됨)
SIN(X)	sine 함수 값. X의 값은 radian임
COS(X)	cosine 함수 값. X의 값은 radian임
TAN(X)	tangent 함수 값. X의 값은 radian임

SAS 기초

비교연산자

연산자	연산 내용
= 또는 EQ	같다.
^= 또는 NE	같지 않다.
> 또는 GT	크다.
>= 또는 GE	같거나 크다.
< 또는 LT	작다.
<= 또는 LE	같거나 작다.
AND 또는 &	그리고
OR 또는	또는
^ 또는 NOT	부정. 아님.

SAS 기초

조건문 IF-THEN/ELSE

IF 문은 IF 다음의 조건이 참일 때 THEN 뒤에 주어진 명령을 실행하고 그렇지 않은 경우에는 ELSE 뒤에 주어진 명령을 실행하는 문장이다.

예를 들어

```
IF YEAR > 1995 THEN MODEL=1;  
ELSE MODEL=0;
```

문의 결과는 변수 YEAR 값이 1995보다 크면 MODEL이라는 변수값을 1로 하고 그렇지 않으면 MODEL 값이 0이 된다.

IF 문의 조건이 참일 때 수행할 명령이 여러 개 있는 경우에는 뒤에서 설명할 DO 문을 이용하면 된다.

ELSE 문은 생략될 수 있으며 그런 경우에는 IF 문에서 준 조건이 맞지 않는 경우에는 아무 일도 일어나지 않는다. THEN 문장 마저 생략될 경우에는 SAS 자료의 부분 집합을 취하고자 할 때 사용되는데 예를 들면 다음과 같다.

SAS 기초

```
DATA JUNIOR;
```

```
SET POP;
```

```
IF AGE <=18;
```

위의 예는 POP라는 SAS 데이터 세트에서 나이가 18세 이하인 경우만 뽑아서 JUNIOR라는 새 데이터 세트를 만든 것이다.

DO 명령

DO 명령은 END 명령이 나올 때까지의 명령을 반복하라는 명령이다. 예를 들어

```
IF X <= Y THEN DO;
```

```
    X = X**2;
```

```
    Y = Y + 3;
```

```
END;
```

라고 하면 X의 값이 Y의 값보다 같거나 작으면 X는 제곱을 하고 Y에는 3을 더해준다.

SAS 기초

DO 문은 반복문(loop)로서 사용할 수도 있는데 이 때는 DO 다음에 루프의 조건을 준다. 예를 들어

```
DO X = 1 TO 100;  
  OUPUT;  
END;
```

라고 하면 지금 만드는 SAS 자료에 X라는 변수가 만들어지고 길이는 100이며 값은 1부터 100이 된다. 이런 루프에서 증가시킬 값이 다르면

```
DO X = 1 TO 100 BY 2;
```

증분이 값이 다르게 줄 수 있다. 이 때 X의 값은 1,3,5,..., 99가 된다. 루프의 증분에 특별한 규칙이 없는 경우에는 해당 값을 직접 줄 수도 있다.

SAS 기초

```
DO X = 1, 5, 8, 11, 57;
```

이라고 하면 X의 값을 1, 5, 8, 11, 57로 바꾸면서 END; 문이 나올 때까지 반복을 하게 된다.

OUTPUT 명령

OUTPUT 명령은 SAS 자료에 현재의 값을 변수값으로 갖는 변수를 만드는 명령이다. 예를 들면

```
/* outputtest1.sas */  
DATA A;  
  DO X = 1 TO 10;  
    IF X ^> 5 THEN Y = 0;  
    ELSE Y = 1;  
    OUTPUT;  
  END;
```