



# 실험계획과 분석

심송용(한림대학교 데이터과학스쿨)

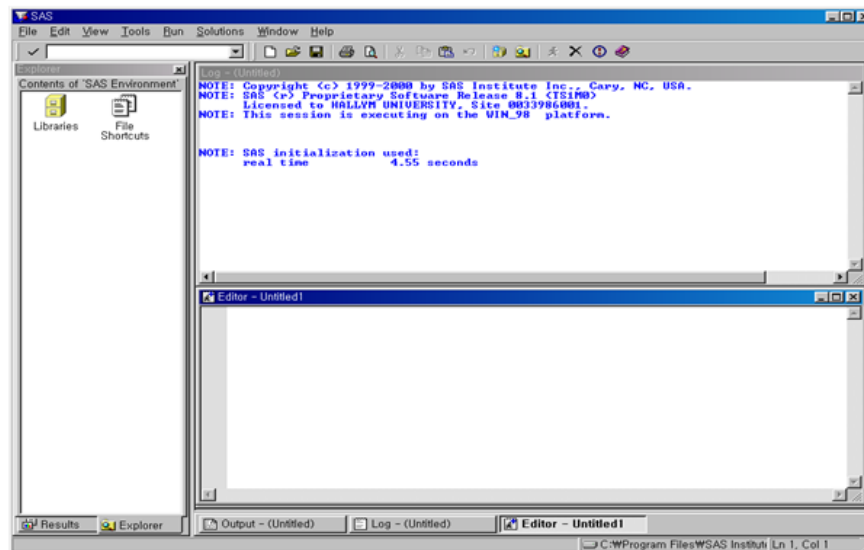
<http://jupiter.hallym.ac.kr>

# SAS 기초

## SAS 명령어

- 한 문장의 명령이 끝나면 세미콜론(;)으로 명령의 끝을 표시
- 하나의 명령문을 여러 줄에 걸쳐 쓸 수도 있고 한 줄에 여러 명령문을 사용할 수도 있음
- 대소문자 구별 없음. 즉 DATA 와 data는 같음

## SAS 일반 버전



# SAS 기초

## SAS 데이터세트(SAS dataset):

- SAS 데이터세트는 SAS에서 개별적으로 이름을 설정하여 사용할 데이터의 모음
- R의 데이터 프레임, 엑셀의 시트
- DATA 문에 의해서 만들어짐.

## DATA 문

- 데이터 세트를 선언하며 이름을 설정한다.
- 데이터 세트의 이름은 특수문자를 제외한 alphanumeric
- .(마침표)를 기준으로 데이터베이스와 테이블로 나누어 설정 가능

## DATA A;

- 데이터 세트 WORK.A를 선언
- 마침표 없으면 WORK가 데이터베이스이름으로 기본 설정
- DATA B.DATA; 등으로 두 개이상의 데이터 세트가 있고 일부 데이터 세트는 이름에 마침표가 있을 시 마침표 없는 데이터 세트는 WORK.을 앞에 붙여야.

## SAS 기초

- KEEP: 뒤에서 읽어들이는 변수 중에서 지금 만들어지는 자료에 포함할 변수의 이름을 준다.
- DROP: 뒤에서 읽어들이는 변수 중에서 지금 만들어지는 자료에 포함하지 않을 변수의 이름을 준다.
- LABEL: 만드는 데이터 세트(SAS dataset)의 설명을 설정한다.
- TYPE 만일 자료가 아주 특이한 형태, 즉 분산-공분산 행렬 등일 때 사용한다.

DATA A(DROP= MON X1-X10 LABEL=A SUBSET OF THE DATA);

- WORK.A라는 자료가 만들어지며 이 자료의 이름은
- LABEL에 주어진 'A SUBSET OF THE DATA'가 된다.
- 뒤에서 INPUT, INFILE 명령 등에서 읽은 변수 중 MON, X1부터 X10은 저장하지 않는다.
- SAS에서 연속적으로 번호가 붙은 변수는 -(마이너스 부호)를 사용하여 한 번에 설정가능한다. 위의 경우 X1-X10은 X1, X2, ... , X10을 의미

# SAS 기초

데이터 세트는 한 번에 여러 개를 만들 수 있다. 예를 들어

```
DATA PAST(KEEP=YR1998) CURRENT(DROP=YR1998);
```

이면 뒤에서 읽을 자료 중에서

- 변수 YR1998은 PAST라는 데이터 세트에,
- 나머지 변수는 CURRENT라는 데이터 세트에 저장된다.
- 이 경우 나중에 YR1998을 포함한 변수들이 SET 명령이나 INPUT 명령 등으로 읽혀질 예정이다.

## 변수의 변환 및 새 변수 계산

변수의 변환은 기존의 변수를 이용하여 새로운 변수를 만들어 내는 것을 말하는 것으로, 예를 들어 키와 몸무게를 이용하여 체질량지수(Body mass index; BMI)를(이때 몸무게는 kg 단위로, 키는 m 단위로 얻는다.)

# SAS 기초

$$\text{BMI} = \frac{\text{몸 무게}}{\text{키}^2}$$

로 계산하려면

$$\text{BMI} = \text{WT}/\text{HT}^{**2};$$

와 같이 연산을 정의하면 된다. 만일 키가 cm 단위로 측정되었으면

$$\text{BMI} = \text{WT}/(\text{HT}/100)^{**2};$$

으로 계산할 수 있다. 각종 연산에서 사용하는 SAS에서 산술 연산자나 내장함수는 뒤에서 자세히 다루기로 한다. 참고로 BMI는 **Quetlet's index**로 불리기도 한다.

## 변수의 선언

SAS 데이터 세트의 변수 선언 (및 읽기)는 INPUT 명령을 사용한다. 이 명령은

```
INPUT var_list;
```

## SAS 기초

으로 사용하며 읽어 들일 변수의 이름을 var\_list에 설정하는 방식으로 사용한다. 이 방식으로 사용하려면 데이터의 값들은 최소 한 개이상의 빈칸으로 구분되어 있어야 한다.

- 변수 이름의 기본은 alphanumeric, 빈칸 및 특수문자 사용불가이다.
- 읽어들일 변수의 값이 문자열이면 변수 이름 뒤에 \$를 붙인다.
- 이 \$는 나중에 사용할 SAS 명령에서는 사용하지 않는다.

예를 들어

```
INPUT YIELD LOC $ X1;
```

은 숫자로 된 변수 YIELD와 X1 및 문자열인 변수 LOC를 읽겠다는 뜻이다.

**결측치:**

SAS 자료에서 결측치는 .(마침표)으로 표시한다. SPSS와 같은 방식이나 R과는 다른 방식이다. 점(.)이 아닌 특정한 자료의 값을 결측치로 설정하고자 하면 MISSING 문을

# SAS 기초

이용하여 다른 값에 대해서도 결측치로 정할 수 있다. 예를 들어

```
/* inputest1.sas */
```

```
DATA A;
```

```
MISSING A V;
```

```
INPUT X Y @@;
```

```
CARDS;
```

```
1 2 2 A 3 V 4 5 5 7 . 6
```

```
;
```

```
PROC FREQ;
```

```
TABLE X Y;
```

```
RUN;
```

로 할 경우에는 Y 값이 A와 V인 경우에는 결측치로  
취급한다. 위 프로그램의 결과는

		Cumulative	Cumulative	
X	Frequency	Percent	Frequency	Percent



# SAS 기초

1	1	20.0	1	20.0
2	1	20.0	2	40.0
3	1	20.0	3	60.0
4	1	20.0	4	80.0
5	1	20.0	5	100.0

Frequency Missing = 1

Y	Cumulative		Cumulative	
	Frequency	Percent	Frequency	Percent
2	1	25.0	1	25.0
5	1	25.0	2	50.0
6	1	25.0	3	75.0

# 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