

Introduction to SAS

SAS (Statistical Analysis System)

- 1966년도에 미국 노스캐롤라이나 대학교에서 농업 연구 과제로부터 시작
- 1976년도에 SAS Institute Inc. 설립
- 1997년부터 매년 가장 일하기 좋은 기업으로 선정 (Fortune지)
- 147개국 75000곳에서 이용

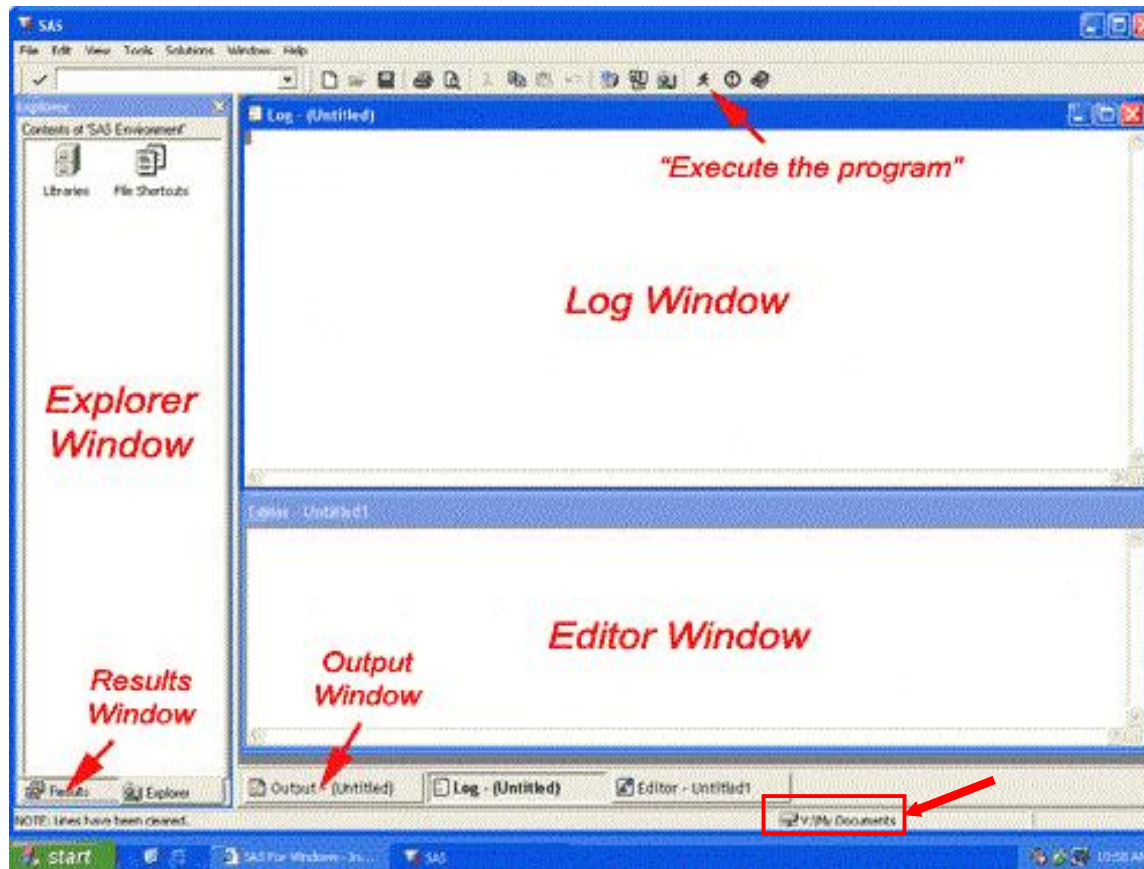
Why SAS?

- 긴 역사, 매우 다양한 프로시저
- 취업시 유용
- 신뢰성, 안정성
- 빅데이터
- 문서작업
- 도움말



1. SAS 윈도우의 구성

- ① 확장편집기 윈도우: SAS 프로그램 작성
- ② 출력 윈도우: 결과출력
- ③ 로그 윈도우:수행과정 및 ERROR 출력
- ④ 기타 (그래프 등의 출력결과, 탐색기)



working directory로 먼저 변경하고 시작

2. SAS 프로그램의 기본구조

DATA step - dataset 생성 및 수정

PROC step - dataset을 가지고 특정 작업 수행

```
/* 기본적인 프로그램의 구조 */  
  
DATA data_name;  
    INPUT  variable1 variable2 $ @@;  
    CARDS;  
        data1 data2 data3 data4 data5 data6  
        data7 data8 data9 data10 data11 data12  
    ;  
  
RUN;  
  
PROC procedure_name OPTION;  
    VAR  variable1;  
    BY variable2; /*BY는 proc sort로 정렬한 변수만 사용*/  
    OUTPUT OUT=data_name_1 keywords; /*저장장소*/  
  
RUN;
```

- 모든 명령문은 세미콜론(;)으로 끝냄
- 대소문자구별 없음
- 설명문 /*...*/ 또는 *.....;
- 한줄에 여러 문장 가능

예제 프로그램

```
/****** example1.sas *****/  
title "Example: getting started with SAS";  
  
data grade;  
    input subject gender$ exam1 exam2 hwgrade$;  
    datalines; /* cards;*/  
10 M 80 84 A  
7 . 85 89 A  
4 F 90 . B  
20 M 82 85 B  
25 F 94 94 A  
14 F 88 84 C  
;  
run;  
  
proc print data=grade;  
    var subject gender;  
run;
```

- 실행 – 마우스로 실행시키고자하는 부분을 블록으로 선택한 후
오른쪽 마우스 클릭하고 submit selection을 클릭 또는 F3 클릭

4. Data step : **dataset** 생성

- SAS editor 에서 **직접 입력** 하거나
- raw data 파일 (예 .txt, .csv, .xlsx, .xls, .dat, .sas7bdat) 을 읽어들이어서(**import**) 생성
=> 확장자 '.sas7bdat' 로 생성됨

1) **직접 입력** : **INPUT** / **DATALINES** or **CARDS**

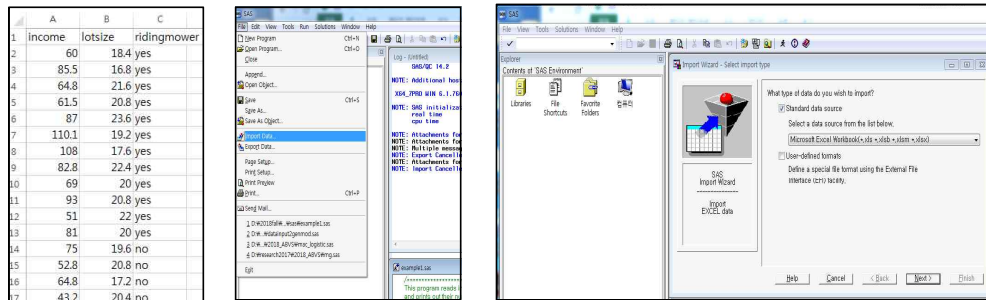
- **DATA** : 데이터셋 이름 입력
- **INPUT** : 변수명 입력
- **DATALINES** or **CARDS** : 자료 입력 시작
- **;** : 자료 끝
- **RUN** : 자료 입력 수행

```
DATA grade;  
  INPUT subject gender$ exam1 exam2 hwgrade$;  
                                     /* gender 와 hwgrade은 문자변수*/  
DATALINES; /* CARDS;로 대체 가능*/  
.....  
;  
RUN;
```

2) 외부 파일을 SAS로 읽기

- File 메뉴 아래의 import data 선택하여 파일 (xls,xlsx, csv, txt 등) 읽기

예) mower.xlsx 자료를 sas로 불러들여 mower 라는 dataset 만들기



참고) 영구적인 DATA 생성

- 위에서 DATA 문으로 생성된 자료는 SAS를 끝내면 사라지는 임시 data set, 즉 work 라이브러리에 저장되어 있음
- SAS를 끝내도 계속 남아 있는 data set 생성을 위해 library를 만들어 저장.

```
LIBNAME sclass 'C:\Users\USER\Desktop\SAS_class\data';  
  
DATA sclass.x;  
    INPUT x1-x3; /*변수이름 x1 x2 x3 */  
    DATALINES;  
        1 1 12.4  
        1 2 11.3  
        1 3 1.4  
        2 1 2.1  
        2 2 19.4  
        2 3 10.0  
    ;  
  
RUN;  
  
proc print data=sclass.x; run;
```

라이브러리 지정

- SAS Library : SAS dataset 이 저장 되는 위치, 컴퓨터 내의 폴더
- Default 라이브러리 : Work library (임시 라이브러리) - SAS 를 나가면 사라지는 임시 자료
- 지정 방법
 - LIBNAME으로 직접 지정
 - Tools -> New Library -> 경로 지정
 - Explorer -> Libraries -> (우 클릭) New -> 경로 지정

3) dataset 다루기 : merge, set, if, label...

- **set** : 포함된 변수들이 동일한 여러 데이터셋의 결합
- **merge**
 - 한 개 이상의 동일 변수를 포함하고 있는 여러 데이터셋을 하나의 동일 변수를 기준으로 결합
 - 각 데이터셋은 결합변수를 기준으로 정렬되어 있어야 함
- 한 데이터셋을 여러 데이터셋으로 나누기

```
data mow_y;  
  set mower;  
  if ridingmower="yes"; *or where ridingmower="yes";  
run;  
data mow_n;  
  set mower;  
  if ridingmower ne "yes";  
run;  
/* 참고*****  
data mow_y1 mow_n1;  
  set mower;  
  if ridingmower="yes" then output mow_y1;  
  else output mow_n1;  
run;  
*/
```



```

data mow_yn;
  set mow_y mow_n;
  label ridingmower="Do they have a riding mower?";
run;

/* 참고 ***** */

proc append base=mow_y data=mow_n;      run;*/

data mow_inc;
  set mower (keep=income ridingmower);
  id=_n_;
run;

/* 참고 ***** */

proc sort data=mow_inc; by id;  run; */

data mow_lot;
  set mower (keep=lotsize ridingmower);
  id=_n_;
run;

data mow_inclot;
  merge mow_inc mow_lot;
  by id;
run;

```

5. PROC step

```
proc contents data=mower; run; /*mower data의 구조와 변수 형식 파악*/  
proc print data=mower (obs=10); /*처음 10개의 observation만 출력*/  
run;
```

```
proc format;  
  value incgrp low-40="Low"  
              40-80="Middle"  
              80<-high="High";  
  value $yngrp "yes"="Y"  
              "no"="N";  
run;  
  
proc print data=mower;  
  format income incgrp. ridingmower $yngrp.;  
run;  
  
data mower_f;  
  set mower;  
  format income incgrp. ridingmower $yngrp.;  
run;  
  
proc means data=mower_f;  
  var income;  
run;  
  
proc freq data=mower_f;  
  table income ridingmower income*ridingmower;  
run;  
  
proc sort data=mower out=mower_s; /* mower data를 income 기준으로 정렬해서 mower_s로 저장*/  
  by income;  
run;
```

6. 결과 저장

- `ods html (pdf, rtf); ... ods html (pdf, rtf) close;`
 - html, xls, pdf, word로 저장
 - html은 default

```
ods pdf file='mower_mean.pdf';
proc means data=mower;
    var income lotsize;
run;
ods pdf close;
```

- proc 실행하면 기본적으로 많은 표들이 출력되는데, 이들 중 특정 표만 출력하고 싶다면

```
ods select <output table 이름>;
```

(각 표의 이름은 "sas proc 프로시저 이름 ods table names"로 구글링)

- 여기에 그 table을 dataset으로 저장하고 싶다면

```
ods output <output table 이름=데이터셋 이름 지정>;
```

- proc report를 이용하여 customized table 출력

```
proc univariate data=mower;
    var income lotsize;
    ods output BasicMeasures=dmeas;
run;
proc contents data=dmeas; run;

PROC REPORT data= dmeas NOWINDOWS HEADLINE;

    title '기초 통계량';
    column VarName LocMeasure LocValue;
    define VarName /width=6 "변수/이름" center;
    define LocMeasure/width=6 center;
    define LocValue/ FORMAT = comma5.2 right;
RUN;
```