

통계계산소프트웨어(2018. 2학기)

# Introduction / SAS프로그래밍 소개

2018. 9. 3.

# 교재 소개

구분	서명	저자	출판사	연도
교재	예제로 배우는 SAS 프로그래밍 입문	김기영 외 3인	자유아카데미	2014
교재	R 컴퓨팅	장영재 외 3인	KNOU PRESS	2015
교재	IBM/SPSS Statistics 기초통계분석	오종철	자유아카데미	2015
교재	Excel 2013	하숙정	한빛아카데미	2015

# 과정 소개

통계계산소프트웨어

**SAS**

**SPSS**

**R**

**엑셀**

# 통계분석소프트웨어

통계 분석을 하려면 먼저 어떤 통계 분석 소프트웨어를 이용할 것인지 선택해야 한다.

최근에는 분야에 따라서 R, SAS, SPSS, STATA, MatLab, S-plus, Minitab, MedCalc, PASS, RevMan, Comprehensive Meta-Analysis, Python, nQuary 등 다양한 소프트웨어가 사용되고 있다.

자료 관리부터 통계분석까지 일련의 과정이 모두 가능한 소프트웨어도 있고, 특정 통계 분석 방법에 특화된 소프트웨어도 있다.

대표적인 몇 가지 소프트웨어의 특성을 비교해보고 자신에게 적합한 소프트웨어를 선택하는 것이 현명할 것이다.

# 통계분석 소프트웨어(통계패키지)

## 1. SAS

- 범용 통계 패키지로 각종 응용프로그램과의 호환이 뛰어남
- SAS시스템은 30여 개의 모듈(소프트웨어)로 이루어져 있음

SAS는 상당히 고가임에도 불구하고 많이 사용되고 있는 소프트웨어이다. 기능이 매우 다양하고 결과에 대한 신뢰도가 높아 여러 공공기관, 사기업, 연구소, 교육기관 등에서 구입하여 사용하고 있다.  
(최근 건강보험공단자료나 심사평가원 자료도 모두 기본적으로 SAS 데이터 셋으로 제공되고 있다.)

SAS도 R과 마찬가지로 기본적인 프로그래밍 언어를 공부해야 한다. 그렇지만 익숙해진 후에는 강력한 통계분석 툴로써 대용량의 자료를 자유자재로 핸들링할 수 있고, 다양한 통계 분석을 수행할 수 있다.

# 통계분석 소프트웨어(통계패키지)

## 2. SPSS

- 광범위한 분야에 대한 데이터 입력, 데이터 관리 및 통계분석을 목적으로 개발
- 전문 프로그래머가 아니라 일반 사용자도 쉽게 사용할 수 있음

SPSS는 유료 통계 분석 소프트웨어지만, 초보자가 사용하기에 좋은 프로그램이다.

통계 분석 방법들이 모두 메뉴(탭)로 구현되어 있어서 몇 번의 클릭만으로 손쉽게 분석할 수 있다.

프로그래밍을 하지 않아도 되므로 사용이 매우 편리한 소프트웨어라고 할 수 있다.

하지만 자료 핸들링이 자유롭지 않기 때문에 새로운 변수를 많이 만들어야 되거나 여러 개의 자료를 이용해서 분석해야 할 때는 다소 답답할 수 있다.

SPSS에서도 명령문을 이용한다면 프로그래밍을 할 수는 있지만 SAS나 R에 비해 자료를 핸들링할 수 있는 범위가 제한적이다.

# 통계분석 소프트웨어(통계패키지)

## 3. R

R의 가장 큰 장점은 '무료'라는 것이다.  
대부분의 통계 분석 소프트웨어가 유료인 반면 R은 오픈소스(open source)로 개인이든 기업이든 모두 무료로 사용이 가능하다.

자유롭게 자료를 핸들링할 수 있고, 다양한 통계 분석 패키지 사용이 가능하다.  
특히 그래프의 경우 다른 소프트웨어에서는 이미 기본적인 형태가 갖추어져 있어서 수정이 어렵지만, 내가 원하는데로 작성할 수 있다는 것이 큰 장점이다.

다른 소프트웨어에서는 그래프의 기본적인 포맷이 정해져 있어 수정이 어려운 경우가 많지만 R에서는 얼마든지 자유롭게 그래프를 만들 수가 있다.  
처음에는 직접 프로그래밍을 해야 한다는 것이 어렵게 느껴질 수도 있지만, 기초적인 내용들만 공부한다면 금방 적응할 수 있을 것이다.

R 다운로드 주소: <https://cran.r-project.org/>

RStudio 다운로드 주소:

<https://www.rstudio.com/products/rstudio/download/>

# 통계분석 소프트웨어(통계패키지)

## 4. 엑셀

미국의 컴퓨터 [소프트웨어](#) 회사인 [마이크로소프트](#)사에서 개발한 윈도 환경의 스프레드시트 프로그램이다.

스프레드시트란 여러 가지 도표 형태의 양식에 계산, 표기되는 사무업무를 자동으로 하는 표계산 프로그램으로 계산기와 계산용지 등이 통합되어 연산 및 표를 작성하고 그래프를 그리는 [소프트웨어](#)를 말한다. 윈도 환경에서 많은 스프레드시트를 연결하고 통합하여 여러 가지 도형과 그래프 등을 작성할 수 있다.

엑셀은 윈도 환경에서 사용자의 그래픽 환경을 제공하는데 스프레드시트 기능을 비롯해 매크로, 그래픽, [데이터베이스](#) 기능과 지도·차트 작성 등 통합 문서작성에 필요한 기능도 제공한다.

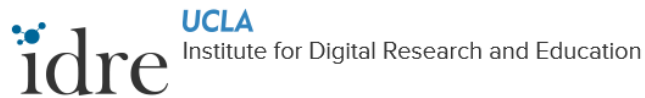
표계산과 그래프, 매크로 등을 하나의 [소프트웨어](#)에서 다룰 수 있는 최초의 통합형 표계산 [소프트웨어](#)이다. 1985년에 초기 버전을 개발한 뒤 잇따라 성능을 향상하여 정밀하며 최고의 기능을 지닌 통합 문서작성 도구이다.

[네이버 지식백과] [엑셀](#) [Excel] (두산백과)



# 통계분석 소프트웨어(통계패키지)

<https://stats.idre.ucla.edu/other/mult-pkg/whatstat/>



Search this website ...

HOME SOFTWARE ▼ RESOURCES ▼ SERVICES ▼ ABOUT US DONATE

## CHOOSING THE CORRECT STATISTICAL TEST IN SAS, STATA, SPSS AND R

The following table shows general guidelines for choosing a statistical analysis. We emphasize that these are general guidelines and should not be construed as hard and fast rules. Usually your data could be analyzed in multiple ways, each of which could yield legitimate answers. The table below covers a number of common analyses and helps you choose among them based on the number of dependent variables (sometimes referred to as outcome variables), the nature of your independent variables (sometimes referred to as predictors). You also want to consider the nature of your dependent variable, namely whether it is an interval variable, ordinal or categorical variable, and whether it is normally distributed (see [What is the difference between categorical, ordinal and interval variables?](#) for more information on this). The table then shows one or more statistical tests commonly used given these types of variables (but not necessarily the only type of test that could be used) and links showing how to do such tests using SAS, Stata and SPSS.

Number of Dependent Variables	Nature of Independent Variables	<a href="#">Nature of Dependent Variable(s)</a>	Test(s)	How to SAS	How to Stata	How to SPSS	How to R
1	0 IVs (1 population)	interval & normal	one-sample t-test	<a href="#">SAS</a>	<a href="#">Stata</a>	<a href="#">SPSS</a>	<a href="#">R</a>
		ordinal or interval	one-sample median	<a href="#">SAS</a>	<a href="#">Stata</a>	<a href="#">SPSS</a>	<a href="#">R</a>
		categorical (2 categories)	binomial test	<a href="#">SAS</a>	<a href="#">Stata</a>	<a href="#">SPSS</a>	<a href="#">R</a>

## 프롤로그

### 데이터 과학자들이 하는 일

데이터 과학자는 통계 및 모델링에 대한 지식을 활용하여 제품 개발에서 고객 유지, 새로운 비즈니스 기회에 이르기까지 모든 것에 대한 데이터를 실용적인 통찰력으로 변환합니다

- 미국에서의 심층 분석 인재에 대한 수요는 2018 년까지 공급보다 50-60 % 증가 할 것으로 예상되며, 14만에서 19만 명의 부족과 150 만 명의 관리자 및 분석가가 부족한 것으로 나타납니다.
- 데이터 과학자는 “21세기의 가장 섹시한 직업“이라고 선언했습니다.
- 2015년 1사분기 동안 데이터 과학자의 취업자 수는 전년 대비 57 % 증가했으며 데이터 과학자 검색은 같은 기간 동안 73.5 % 증가했습니다.
- 데이터 과학자는 높은 수입 잠재력, 풍부한 경력 기회 및 일자리 수 때문에 **미국에서 최고의 직업으로 선정**되었습니다.

## 프롤로그

### 성공을 위해 필요한 기술

데이터 과학자들은 기술, 분석 및 프리젠테이션 기술의 고유 한 조합을 보유하고 있어야 하기 때문에 찾기 어렵습니다.

그들은 **통계와 응용 수학을 이해**합니다. 그들은 설계 한 실험으로 가설을 검증 할 수 있습니다. 이들은 취합, 처리 및 데이터 저장 방법을 설계 할 수 있는 충분한 **프로그래밍 지식**을 갖추고 있습니다. 또한 **데이터 시각화 및 스토리텔링을 통해 결과를 전달**합니다.

그들이 사용하는 언어 및 응용 프로그램 중 일부는 **SQL, R, Python, SPSS, Tableau 및 Hadoop**입니다.

# 통계와 통계학

## □ 통계(statistic)

- 사회적 현상 혹은 자연 현상 규명하기 위해 수집된 각종 데이터 요약하거나 적절한 방법 통하여 일차적으로 가공되어 나오는 정보
- 숫자, 그래프, 도표, 그림 같은 데이터로부터 얻어진 초보적 정보
  - 숫자로 요약되어 저장, 유통되는 숫자정보
  - 정리되어서 발표되는 일종의 정리된 숫자정보

### [숫자정보 예]

분야	숫자정보(통계)
날씨	비율 확률, 기온, 불쾌지수, 빨래지수
경제	GDP, 인플레이션율, 주가지수, 환율, 금리, 실업률
여론조사	대통령 선거 관련 여론조사, TV 시청률 조사
마케팅	소비자 만족도 조사, 신상품 수요조사
스포츠	타율, 장타율, 점유율 등
교육	IQ, 수학평균점수 등

# 통계와 통계학

## □ 통계학(statistics)

- 자료를 과학적으로 수집하고 적절한 방법으로 정리, 분석하는 학문
  - 현실에서 정확한 대상을 선정하고, 연구목적에 필요한 자료를 경제성과 정밀도를 고려하여 최적의 방법으로 수집하고, 수집된 자료가 과학적인 이론에 의하여 정리, 분석되도록 방법을 제시 해주는 학문
  - 숫자정보를 수집, 정리, 기술, 분석하고 그 결과를 해석하는 규칙과 절차를 연구하는 학문
  - 과학적인 이론과 근거에 의해 불확실한 사실에 대해 과학적인 판단을 내리는 방법 제시
  - 불확실한 상황에서 타당한 결론을 도출하여 최선의 의사결정이 이루어 지도록 하는 목적
- 
- 자료에서 얻은 정보를 이용하여 전체(모집단)에 대한 정보로 바꾸는 과정에서 필연적으로 수반되는 오차에 관한 이론의 정립, 응용 과정에서의 문제점 발견과 해결방법을 연구하는 학문

# 통계와 통계학

## □ 통계학(statistics)



# 통계와 통계학

- 영문으로는 모두 Statistics 이다. 그러나 통계는 통상 정리되어서 주기적으로 발표되는 일종의 정리된 정보인 반면, 통계학은 자료에 대한 학문이다. 자료가 있으면 이를 어떻게 체계적으로 정리하고, 요약하는가에 대한 학문이다.
- 통계를 잘 알려면 해당 분야의 내용을 잘 알아야 하지만 통계학을 알 필요는 없다. 그러나 통계학을 잘 하려면 수학 등 통계적 방법에 대한 이해를 잘해야 한다. 통계학을 이용하여 해당 분야에 대한 내용을 더 잘 알게 되는 것이 최종 목표일 수도 있다. 예를 들어, 경제성장률의 의미를 알기 위해서는 국민계정의 작성과정만 잘 알고 있으면 가능하다. 그러나 국민성장률을 숫자로 예측한다면 국민계정의 작성보다는 예측 관련 통계학을 알아야 하는 이치이다.

(출처 : 통계교육원)

# 과정 소개

## 프로그래밍이란

수식이나 작업을 컴퓨터에 알맞도록 정리해서 순서를 정하고 컴퓨터 특유의 명령코드로 고쳐 쓰는 작업을 총칭해서 프로그래밍이라 하고, **컴퓨터의 명령 코드를 쓰는 작업을 특히 코딩(coding)**이라고도 한다. 컴퓨터가 처음 나타난 1950년대 초기까지는 프로그래밍은 숫자를 나열한 명령코드를 쓰는 것이었다. 이것을 기계어(machine language)라 한다.

그러나 기계어에서는 틀리기 쉽고, 또한 틀린 곳을 발견하기가 어렵다는 등 작업하기가 곤란하므로, 그 후 인간이 외우기 쉬운 기호나 언어·수식을 사용해서 프로그램을 쓰고, 그것을 일단 컴퓨터에 넣어서 컴퓨터 자신의 명령코드로 고쳐 그것으로부터 계산을 실시하는 방식이 고안되었다. 이것은 프로그램을 만드는 작업의 일부를 컴퓨터 자체에 부담시켜 작업 능률을 향상시키자는 방식이다.

[네이버 지식백과] 프로그래밍 [programming]



# 과정 소개

## 프로그래밍 언어

- 컴퓨터 프로그래머의 의사를 전달하는 방법
- 프로그램을 작성하는 형식
- 프로그래머가 컴퓨터를 어떻게 추상화해야 하는가에 영향을 미침
- 컴퓨터가 읽을 수 있고 사람이 읽을 수 있는 형식으로 계산을 서술하는 표기체계

## 프로그램

- 컴퓨터와 사람이 동시에 이해할 수 있는 형식으로 작성된 글



# 프로그래밍 언어

X = 1  
LET X = 1

사용자중심  
스크립트형

SQL, **SAS**, HTML

사용자중심  
객체지향

C(70년대), JAVA(90년대)

사용목적  
절차중심,문제중심

COBOL, FORTRAN  
1950년대

MOV EAX

어셈블러

인터프리터

컴파일러

프로그래밍 언어 구현기법

기계어

0100 0210

하드웨어

# 강의순서

1. SAS 프로그래밍/ SAS 시스템
2. SAS Data Step
3. SAS 데이터 단계에서 사용되는 명령문
4. SAS 기본 프로시저
5. SAS 그래프/ 데이터 요약 및 표현

6. SPSS ( Introduction/자료 불러오기 )
7. SPSS ( 자료의 변환/ 데이터 관리 )

## 8. SAS/SPSS 기초통계분석

9. R 프로그래밍 입문
10. R 데이터 구조/ R 데이터 탐색
11. R 프로그래밍 응용
12. R 응용 : 데이터분석 사례연구

## 13. EXCEL을 이용한 통계분석

# 강의 계획

강의	<ol style="list-style-type: none"> <li>1. 기간 : 2018년 9월 3일 ~ 2018년 12월 12일</li> <li>2. 시간 : 매주 월/수</li> <li>3. 장소 : 정경관 206호</li> <li>4. 구성 : 강의 + 실습</li> </ol>
평가	<ol style="list-style-type: none"> <li>1. 평가 : 출석(10%), 과제(15%), 중간시험(35%), 기말시험(40%)</li> <li>2. 평가등급 : A+, A (~35%), B+, B, C+, F(35 ~ 100%)  <b>** 중간, 기말 불 응시 : F</b></li> <li>3. 시험 : 중간, 기말</li> <li>4. 참여도 : 결석 시 감점(1회 0.2점)</li> </ol>
학습 방법	<ol style="list-style-type: none"> <li>1. 강의자료 다운로드 : Blackboard</li> <li>2. 과제 제출 : 하드카피 제출</li> </ol>
연락처	<p>M. 010-3720-6920 / 통계연구소(정경관 411호)</p> <p>E. oh1207@nate.com</p>



# SAS 프로그래밍

SAS 시스템이 이해할 수 있도록



1. SAS 시스템내에서
2. SAS 프로그래밍언어로
3. SAS의 프로세스를 만드는 것

# 과정 소개

## SAS ?

### Company Facts & Financials

#### Customer

##### Number of Countries Installed

SAS has customers in 147 countries.

##### Total Worldwide Customer Sites

Our software is installed at more than 83,000 business, government and university sites.

##### Fortune Global 500® Customers

96 of the top 100 companies on the 2017 Fortune Global 500® are SAS customers.

#### Employee

##### Worldwide Employees

14,328 total employees

##### Breakdown by Geography

United States: 7,156

World Headquarters (Cary, NC): 5,753

Canada: 328

Latin America: 491

Europe, Middle East and Africa: 3,723

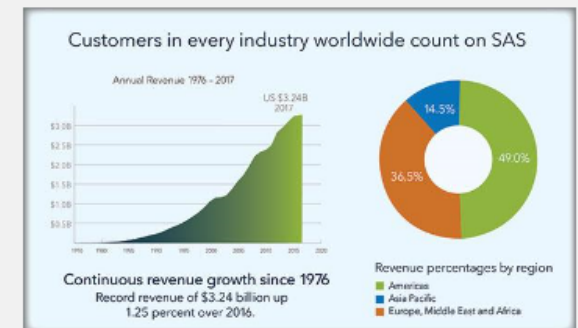
Asia Pacific: 2,630

#### Financial

##### Worldwide Revenue

2017 Revenue: US\$3.24 billion

##### Historical revenue data



# 과정 소개

## Why SAS ?

### (1) 사용의 용이성

- 단 몇 줄의 프로그램으로 간단한 작업 가능하고 에러 유형 파악이 용이
- free format 형식이므로 사용이 간편

### (2) 간단한 통계분석과 폭넓은 통계기법의 제공

- 통계적 이론의 실제적 의미를 쉽게 이해할 수 있음
- 다양한 통계분석 보고서 및 통계 분석라이브러리를 제공

### (3) 자료관리와 처리의 효율성

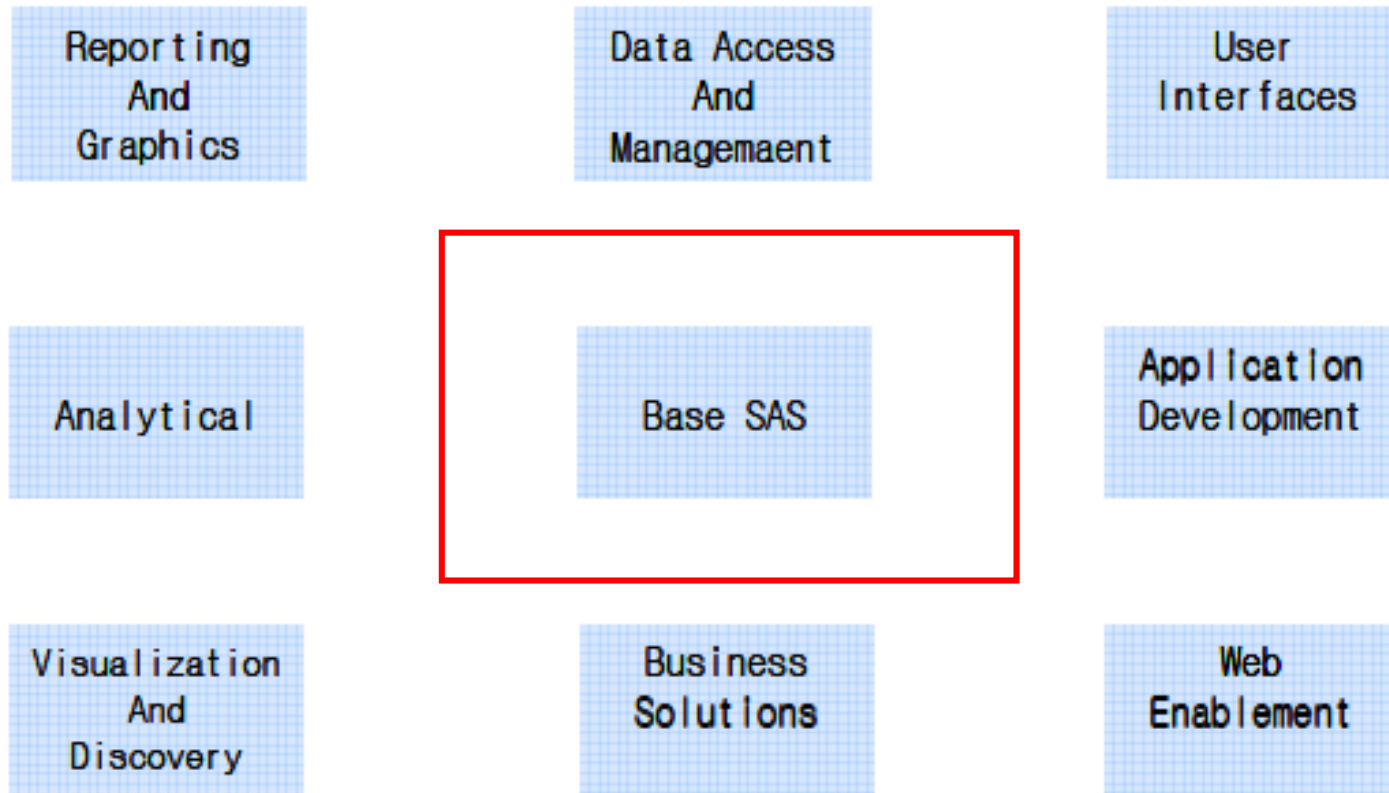
- 데이터의 입력과 편집이 수월
- DBASE나 메모장, Excel등에서 저장된 Data 이용 가능 (호환성)

### (4) 신뢰성 및 범용성

- 가장 높은 비율로 논문 인용, 최대 사용자

# 과정 소개

- SAS 컴포넌트 구성





# SAS 소개

- BASE SAS

**SAS/CORE** : SAS system의 관리와 제어를 담당.

**SAS/BASE** : 자료의 입출력과 저장, SAS file의 편집, 보고서 작성 등을 담당.  
간단한 기초 통계분석도 할 수 있는 SAS system의 기본

- Analytical

**SAS/STAT** : 통계모형을 이용한 통계분석 Procedures모임

**SAS / ETS** : 계량경제, 예측 및 재무관련 자료처리 기능

**SAS / O R** : 프로젝트 관리 및 운용연구(operations research)

**SAS / Q C** : 통계적 품질관리 소프트웨어

**SAS / IML** : 수학, 엔지니어링 등의 행렬 계산

**SAS/MACRO** : 동일한 작업의 반복을 단순화시키는 소프트웨어

- Report & Graphics

**SAS/GRAPH** : 각종 형태의 도표, 지도, 슬라이드 작성, PRINTER, PLOTTER 지원

# SAS 소개

- User Interface

**SAS/INSIGHT** : 대화식 통계 그래픽 기능 .

**SAS Enterprise Guide** : 프로그래밍 없이 메뉴방식으로 지원

- Visualization and Discovery

**SAS Metadata Server , SAS Enterprise Miner**

- Data Access & Management

**SAS/ACCESS** : 다양한 DB포맷의 입출력 지원

- Web Enablement

**SAS/IntrNet** : 웹기반 보고서 출력 및 분석 기능 제공

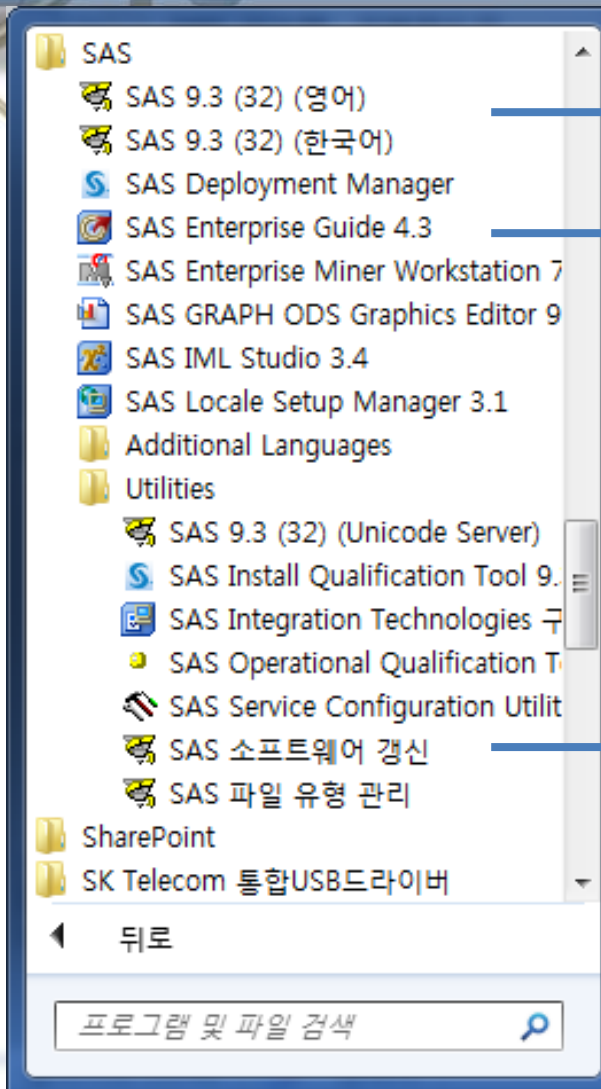
- Application Development

**SAS/AF      SAS / EIS**

- Connect & Share

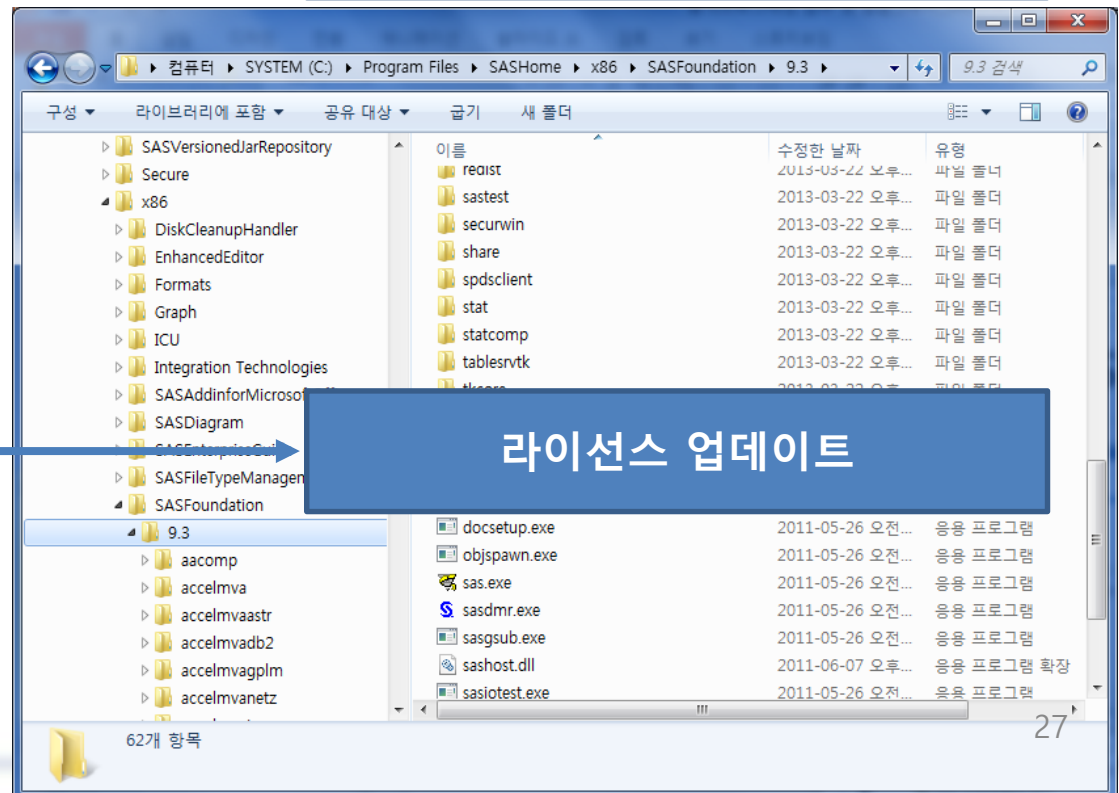
**SAS/Connect      SAS / SHARE**

# SAS 시스템 구동하기



SAS실행아이콘

대화형/프로젝트 관리



라이선스 업데이트

# SAS 시스템 구동하기

◦ 구동 : 시작 > 프로그램(P) > 모든프로그램 > SAS > SAS 9.3 /9.4

The screenshot shows the SAS 9.3/9.4 desktop environment. The interface includes a menu bar at the top, a toolbar, a left sidebar with a file explorer, and several open windows. Red boxes and numbers highlight specific features:

- ① 메뉴 (Menu): Points to the menu bar at the top.
- ② 메뉴별 기능 버튼 (Menu-specific function buttons): Points to the toolbar below the menu bar.
- ③ 탐색기 (Explorer): Points to the left sidebar containing the file explorer.
- ④ 확장편집기 (Extended editor): Points to the '확장 편집기 - 제목 없음1' window.
- ⑤ 로그윈도우 (Log window): Points to the '로그 - (제목없음)' window.
- ⑥ 출력윈도우 (탭) (Output window (tab)): Points to the '출력 - (제목없음)' window.

Additional text in the image includes:

- 탐색기 (Explorer)
- 'SAS 환경'의 내용 (Contents of 'SAS environment')
- 라이브러리 (Library)
- 파일 바로 가기 (File shortcuts)
- 즐거 찾는 폴더 (Folders to enjoy finding)
- 컴퓨터 (Computer)
- 로그 - (제목없음) (Log - (Untitled))
- NOTE: Copyright (c) 2002-2010 by SAS Institute Inc., Cary, NC, USA.
- NOTE: SAS (r) Proprietary Software 9.3 (TS1M0 DBCS3060)
- Licensed to KOREA UNIVERSITY, Site 10503207.
- NOTE: 이 세션은 W32\_ZHOME 플랫폼에서 실행되고 있습니다.
- NOTE: SAS 시스템 초기화 실행:
- 실행 시간 2.63 초 (Execution time 2.63 sec)
- CPU 시간 1.16 초 (CPU time 1.16 sec)
- 확장 편집기 - 제목 없음1 (Extended editor - Untitled1)
- 출력 - (제목없음) (Output - (Untitled))
- 로그 - (제목없음) (Log - (Untitled))
- 확장 편집기 - 제목 없음... (Extended editor - Untitled...)
- C:\Users\Administrator
- Ln 1, Col 1



# SAS 시스템- 메뉴

- 파일 메뉴(F)
  - 프로그램(또는 데이터)을 저장하거나 불러오는 메뉴
- 편집메뉴(E)
  - 실행취소, 복사, 붙여넣기, 지우기 등 프로그램편집기에서의 편집기능과
  - 출력, 로그 윈도우에서의 선택삭제 기능을 제공하는 메뉴
- 도구 메뉴 (T)
  - 테이블(데이터) 편집기, 그래프편집기, 리포트편집기 등의 편집기윈도우 오픈
  - 옵션 설정 - 시스템 옵션 및 출력옵션 지정
- 실행 메뉴 (R )
  - 편집기의 프로그램을 실행시킴, 일부 선택실행 가능함
- 솔루션 메뉴 (S)
  - 대화형으로 제공되는 분석시스템이나 BI 어플리케이션 실행

# SAS 시스템 윈도우

- 확장편집기 윈도우

- ✓ 프로그램을 작성하고 편집하는 윈도우
- ✓ 풀다운 메뉴의 submit 메뉴나 도구상자의 submit 아이콘을 클릭하여 프로그램을 실행한다.

- LOG 윈도우

- ✓ SAS가 명령을 수행한 잡업내용을 보여줌. 프로그램 상 발생한 문법오류, 주의사항, 경고 등을 보여준다.

- 출력 윈도우

- ✓ 프로그램 수행의 결과를 보여주는 윈도우

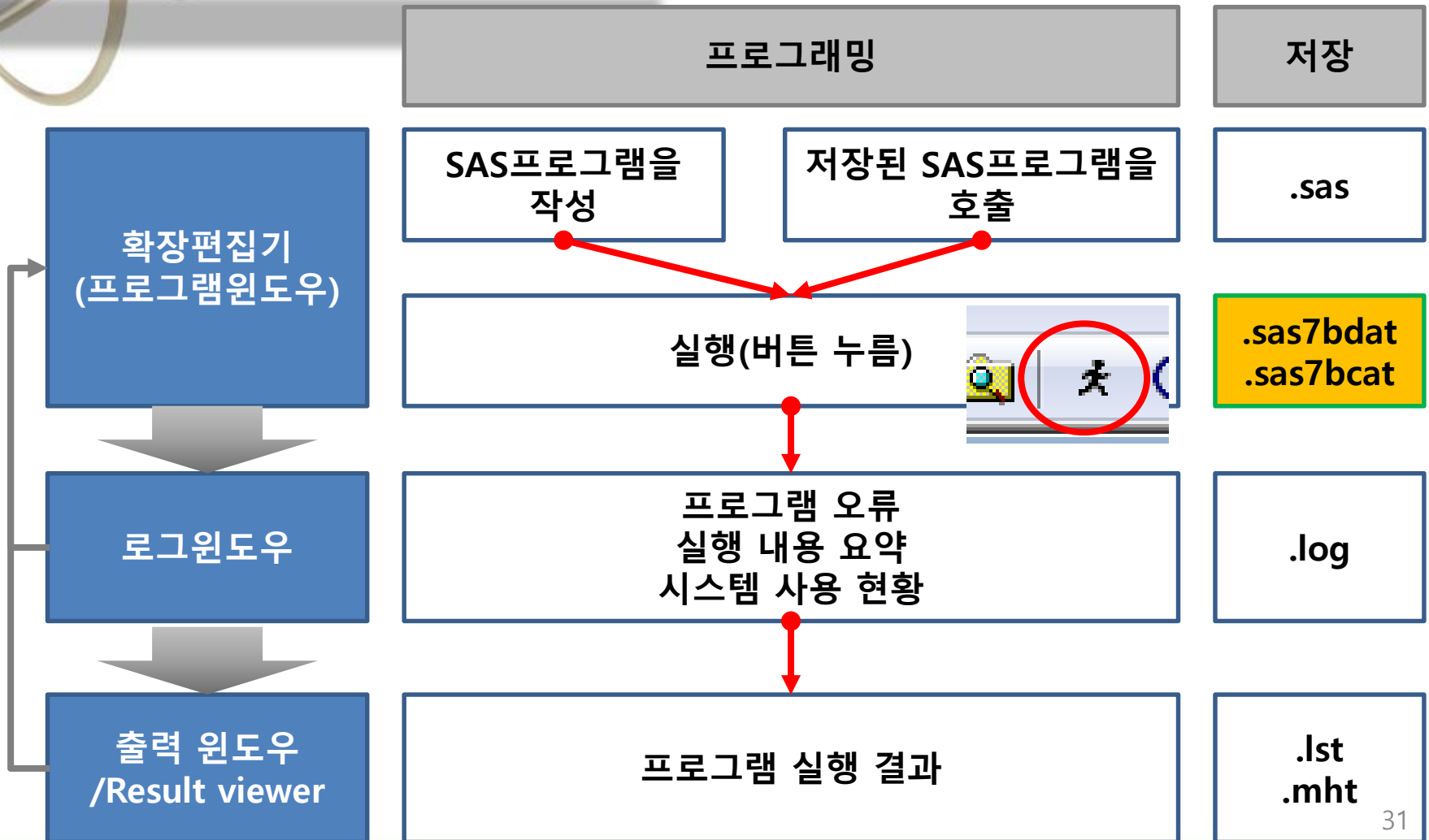
- 분석결과 탐색기

- ✓ OUTPUT 윈도우나 GRAPH 윈도우에서 출력되는 결과를 조회 관리함

- 라이브러리 탐색기

- ✓ SAS 라이브러리와 데이터셋을 관리하는 윈도우

# SAS 프로그램 실행 순서



# SAS 프로그램 - 구성요소

- 토큰(명령어)는 예약어, 명칭 과 기호로 구성되며,  
여러 개의 토큰(명령어) 조합으로 하나의 명령문이 이루어짐

## KEYWORD (예약어)

SAS시스템이  
사전 정의된 명령어

DATA PROC  
RUN  
DO TO END  
OUTPUT

## name (명칭)

SAS 시스템에서  
식별할 수 있도록  
사용자가 만든 이름

age  
i  
test

## SYMBOL (기호)

SAS 프로그램에서  
특수한 역할을 수행하는  
특수문자와 연산자

;  
=  
+ - \* / ( )  
\$ @ # ! & %

## 명령문

Compile을 위한 기본 문장  
하나의 기능을 정의/수행

```
DO i = 1 TO 10 BY 2 ;  
PROC means DATA=test NWAY NOPRINT ;  
ARRAY dif[10] diff01-diff10 ;  
DATA test ;  
RUN ;
```



# SAS 프로그램 - 단계

- 하나의 단계에는 여러 개의 명령문의 조합으로 이루어짐.
- 데이터생성단계는 DATA 키워드로 시작하고 분석단계는 PROC 키워드로 시작하며, 각 단계는 RUN ; 으로 종료됨

