

1-1강

과정소개

1. SAS - 라이선스 비용 지불
2. SPSS - 라이선스 비용 지불
3. R - 무료
4. 엑셀 - 범용

데이터분석 소프트웨어(통계패키지)

1. SAS

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

2. SPSS

- 광범위한 분야에 대한 데이터 입력, 데이터 관리 및 통계분석을 목적으로 개발
엑셀과 굉장히 유사
- 전문 프로그래머가 아니라 일반 사용자도 쉽게 사용할 수 있음
그리고 사용하기 쉽다.

결과를 보고 해석을 해야하는데, 그때는 통계학의 다양한 지식들이 필요하다

3. R

- 가장 큰 장점은 인터넷에서 자유롭게 다운받아 사용할 수 있는 소프트웨어
- 자료처리 및 그래픽스 분야에 탁월
R project.org라는 것이 있다.
패키지가 상황에 따라 stable 하거나 아닌 것도 있다.

4. 엑셀

- 마이크로소프트에 의해 개발된 스프레드시트 프로그램,
 - 스프레드시트란 컴퓨터에서 계산, 자료관리, 차트의 작성, 검색 등을 효율적으로 할 수 있게 하는 응용 프로그램

통계패키지 비교

	SAS	SPSS	R
비용	고가	고가	무료
용량	대용량	대용량	저용량
모듈지원	별도구매	별도구매	무료
최신기술반영	느림	조금 느림	빠름
학습자료	유료도서 위주	유료도서 위주	공개논문 및 자료
공식 질의 커뮤니티	없음	없음	활발함

R - 128GB 이상이면 좀 무리가 있다.

SAS, SPSS는 모듈을 별도 구매를 해야한다.

R은 다 무료

SAS는 1~2년 정도로 바뀐다. (stable하기 때문에)

SPSS는 바뀔 때 혁신적으로 바뀐다.

SAS는 자료가 잘 안나온다.

교재

구분	서명	저자	출판사	연도
교재	예제로 배우는 SAS 프로그래밍 입문	김기영 외 3인	자유아카데미	2014
교재	R 컴퓨팅	장영재 최상범 한승봉	KNOU PRESS	2020
참고	SPSS24를 이용한 통계분석	서의훈	자유아카데미	2018
참고	Excel 2013	하숙정	한빛아카데미	2015

SAS는 블로그에 자료도 많다.

R은 꼭 책을 사라

Excel은 2013은 pdf파일을 따로 받았다.

강의내용

1. SAS 프로그래밍/ SAS 시스템
2. SAS Data Step
3. SAS 데이터 단계 명령문
4. SAS 데이터 셋의 결합 및 관리
5. SAS 데이터 요약 및 표현
6. SPSS 사용법(Introduction/자료 불러오기)
7. SPSS 사용법(자료의 변환/ 데이터 관리)
8. SAS/SPSS 기초통계분석
9. R 프로그래밍 입문
10. R 데이터 구조
11. R 함수/그래프

12. R 데이터탐색/기초통계분석

13. EXCEL 기초

SAS University를 다운 받아라

미국최고의 직업은 "데이터과학자"

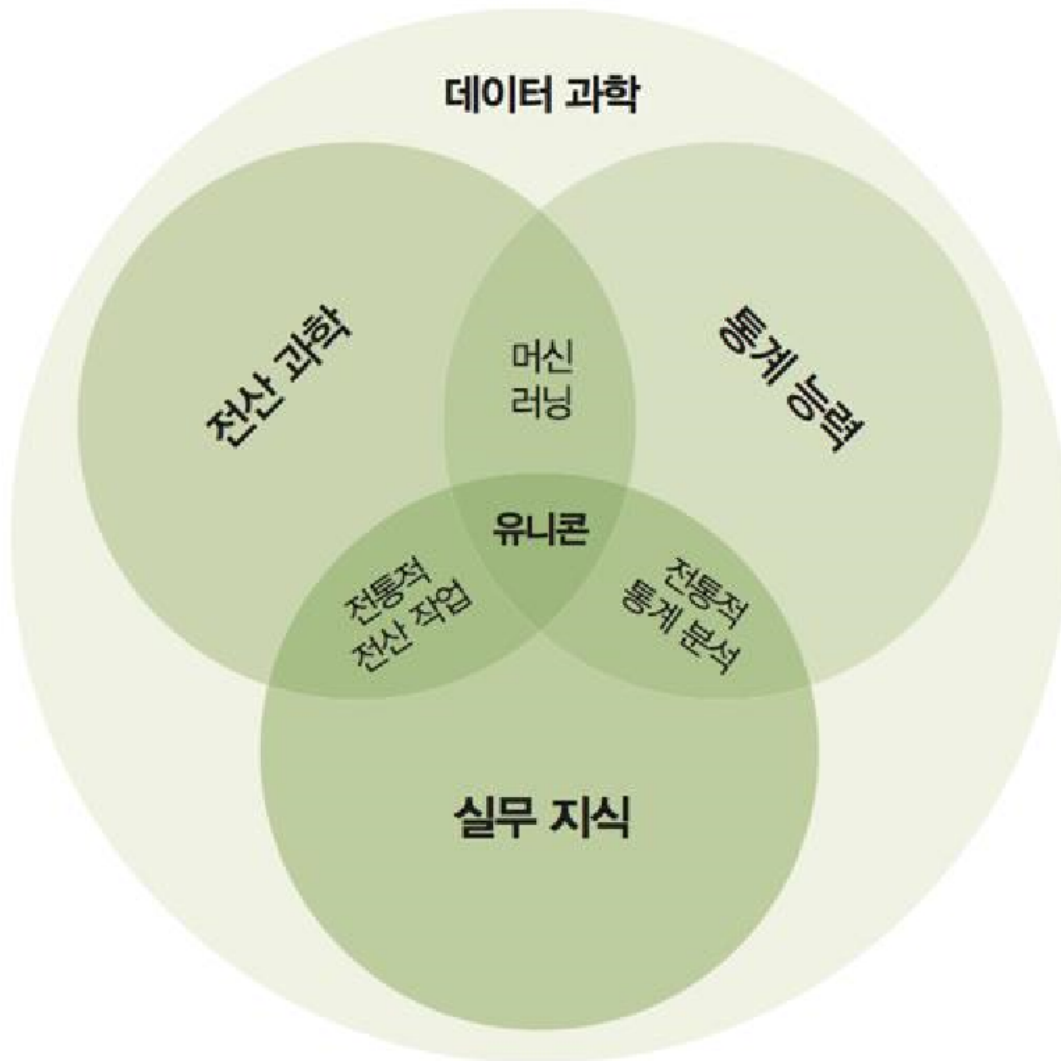
데이터과학자가 갖춰야 할 8가지 역량

- 첫 번째는 수학이다.
- 두 번째는 통계학이다.
- 세 번째는 프로그래밍, 코딩이다.
- 네 번째는 바로 머신러닝이다.
- 다섯째는 하둡(Hadoop)이다.
- 여섯째 시각화다.
시각화에 뛰어난 것이 R이다.
- 일곱째는 분야 전문성이다.
- 여덟째는 커뮤니케이션이다.

데이터과학자가 갖춰야 할 능력은

- ✓ 실제적인 문제를 통계적으로 표현하고,
 - ✓ 컴퓨터 도구를 사용하여 시각화와 데이터 가공과 모형화를 한 후에,
 - ✓ 그를 이용 하여 실제적인 언어로 의미 있는 결과를 만들어내 는 능력의 조합임을 알 수 있다.
- + 협업 능력

데이터과학자가 갖추어야 할 능력



핸드폰과 카드를 보면 그 유동인구를 알 수가 있다.

과정소개

통계(statistics)란

통계학은 표본의 자료를 수집, 정리, 요약하고 나아가 요약된 자료를 토대로 그 자료의 모태가 되는 모집단에 대해 짐작, 추측해 보는 작업을 포함

1. 모집단 현상,이론, 가설
2. 표본조사 실험,설문, 측정

3. 수집/가공 (데이터 정보화, 도식화)

4. 분석/추론 (추정,검정, 예측)

프로그래밍이란

“수식이나 작업을 컴퓨터에 알맞도록 정리해서 순서를 정하고 컴퓨터 특유의 명령코드로 고쳐 쓰는 작업을 총칭”

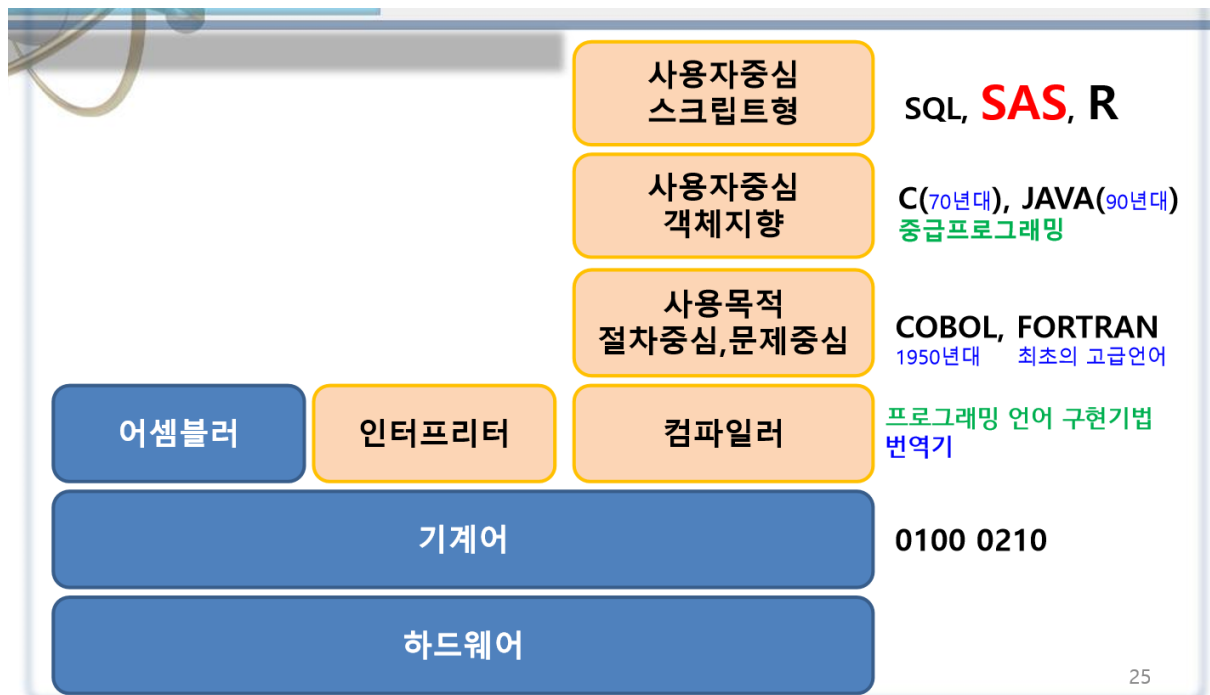
프로그래밍 언어

- 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



4세대 언어가 SAS, R 정도이다.

데이터 정보와 컴퓨터

컴퓨터를 통하여 데이터정보를 효율적으로 처리하려면 다음과 같은 사항이 요구된다

1. 컴퓨터를 잘 활용할 수 있어야 한다
2. 지식정보화 사회는 정보가 중심이 되어 이루어 지는 사회로, 인터넷의 활용이 필수적으로 요구된다. 따라서 인터넷을 효율적으로 활용할 수 있는 능력이 필요하다
3. 정보를 판단하는 능력이 중요하다.
4. 데이터분석에 대한 기본개념과 방법론을 숙지해야 한다.
5. 데이터로 부터 유용한 정보를 추출하고 합리적인 의사결정을 하려면 데이터분석 소프트웨어의 활용이 요구되며, 따라서 이러한 소프트웨어의 활용법을 숙지해야 한다.