

과제1

2018320161_송대선

Q. 통계패키지 SAS, SPSS, R, EXCEL의 장단점을 기술하십시오.

- SAS

매우 크기가 큰 자료를 읽거나 쓰는데 용이하며, 다양한 분석을 시도해야 하는 경우에 유용합니다. 즉, 대용량의 파일을 안정적으로 제공할 수 있다는 것입니다. 파일 사용에 있어서 융통성이 있어서 호환이 쉽고, 운영체제 상관없이 어디에서나 이용할 수 있습니다. 그리고 통계분석의 종류가 많고 강력한 데이터구조 변경능력을 가지고 있어 통계이론의 실제적 의미를 보다 쉽게 파악할 수 있습니다. 그리고 다양한 통계분석 보고서와 강력한 라이브러리들을 제공합니다. free format 형식을 가지고 있어 사용이 상대적으로 편하고, 여러 유형 파악이 용이합니다. DB, 메모장, Excel등에 저장된 데이터도 이용이 가능할 정도로 호환성이 좋습니다. 그리고 무엇보다 굉장히 stable한 통계패키지이기에 신뢰성이 가장 높고, 가장 높은 비유로 논문 인용에 사용될 정도입니다. 관련 자격증 시험도 다양하게 있어서 자신의 능력을 공적으로 증명할 수가 있습니다. R 등의 무료 S/W, 범용성이 뛰어난 Excel 등을 활용하는 추세이지만 실제 통계작성기관의 업무에서는 아직도 SAS가 많이 활용됩니다. 하지만 프로그램의 크기가 크고, 라이선스 비용이 매우 비쌉니다. 1년간 사용할 라이선스를 얻는데 500 - 1,400만원 정도 한다고 합니다. SAS의 SAS는 SAS 전용 프로그래밍 언어를 사용하는데, 1세대, 2세대, 3세대 프로그래밍 언어들에 비해서는 배우기가 쉬우나, SPSS와 같은 통계패키지에 비해서는 어렵습니다. 그리고 stable한 만큼 최신기술에 대한 반영이 느리고 보수적입니다. 공식 질의 커뮤니티가 없고, 공부할만한 자료가 나오지 않습니다.

- SPSS

SPSS는 광범위한 분야에 대한 데이터 입력, 데이터 관리 및 통계분석을 목적으로 하여 개발된 통계프로그램입니다. SPSS는 마우스를 이용한 GUI 인터페이스 위주로 되어 있기 때문에 SAS에 비해서는 조금 더 직관적입니다. 많은 부분에 사용되는 엑셀과도 유사해 보이기에 진입장벽이 낮습니다. 그래서 전문 프로그래머가 아닌 일반 사용자도 쉽게 사용할 수 있습니다. 상경계나 사회과학 관련 분야에서 자주 사용되는 통계패키지입니다. Windows, Mac, Linux, Unix 등 다양한 OS에서 사용가능하다는 장점이 있습니다. 하지만 마우스를 주로 사용하기 때문에 데이터의 양이 너무 많아지면 효율이 잘 안 나오고, 전문적인 작업을 하기가 힘들어집니다. 그리고 무료 통계패키지가 아니기에 라이선스 비용을 지불해야 합니다. 개인용 프로그램 가격이 무려 400~600만원. 기업용은 1000만원이 가볍게 넘어

가며, 공용 네트워크 같은 데 쓰려면 라이선스 비용으로 1억원을 내야 합니다. 그리고 프로그램의 크기 자체도 큰 편입니다.

- R

R의 가장 큰 장점은 무료 통계패키지이기에 인터넷에서 자유롭게 다운로드 받아 사용할 수 있다는 것입니다. 수많은 R 통계패키지가 이미 개발되어 있기에 특수한 경우가 아니라면 굳이 직접 개발해야 할 필요가 없습니다. 통계 전문 언어 중 가장 보편적이기 때문에 내가 사용하고 싶은 모든 통계 기법이 이미 어딘가에 패키지 형태로 구현되어 있다고 봐도 무방할 정도입니다. 그 수많은 R 통계패키지조차 전부 무료이기에 패키지를 설치하면서 무수히 많은 기능들을 추가할 수 있습니다. 공개논문 및 자료들로 학습을 할 수 있으며, 공식 질의 커뮤니티가 활발히 운영되어 궁금한 점이 있으면 답을 얻기도 편합니다. 패키지를 직접 만드는 것에도 열려있어 자유롭습니다. 자료처리 및 그래픽스 분야에 탁월하여 그 결과물의 가독성이 높습니다. 하지만 단점으로는 큰 데이터 집합을 사용할 때 무리가 생길 수 있다는 점이 있습니다. 다른 통계패키지에 비해 상대적으로 처리 속도가 느립니다. 그리고 패키지를 만드는 것에 대해 자유로운 만큼, 때때로 패키지가 unstable할 때도 있습니다.

- 엑셀

매우 범용적인 통계프로그램입니다. 통계와 직접적인 관련이 없는 일반회사들에도 많이 사용 됩니다. 기초적인 자료 분석이나 통계에 능통합니다. 300개가 넘는 다양한 함수들을 지원하며 쉽게 사용할 수 있습니다. 회사 실무에 최적화되어 있습니다. 다양한 기능이 지원되기에 실무에서 원하는 대부분의 기능들을 사용 가능합니다. 시각화와 문서화에 뛰어납니다. 공식자격증이 있으며, 인터넷에서 엑셀 관련 정보, 활용법을 쉽게 얻을 수 있습니다. 하지만 프로그램의 목적 자체가 통계분석만을 위한 것이 아니기에 데이터의 양이 많아지면 속도가 급격하게 느려지며, 아주 복잡한 연산은 무리가 있습니다. 그리고 외부데이터를 가져올 때 자릿수가 너무 큰 수에 대해서는 뒷부분을 0으로 처리하기에 자릿수가 큰 경우 정확한 데이터를 가져오는 것이 어렵습니다.