

AI기초통계

담당교수 : 김 승 재

< 공 지 사 항 >

- 개강 : 9월 01일(수)
- ‘코로나19로 **대면-비대면** 교차수업 진행
<1주차~7주차: **대면**, 8~15주차: **비대면**>
 - 1주차 : 9월 01일 (월) ~ 9월 07일(금)
 - 2주차 : 9월 08일(월) ~ 9월 14일(금)
- **대면-비대면** 수업 참여 방법
 - 대면 : 지정된 강의실 입실(시간준수)
 - 비대면 : 호남대학교 **LMS 시스템** 이용하여 참여
- 수업 및 **대학생활**에 필요한 사항은 별도 **공지**

대면 강의 내용(1주차)

○ 1주차 강의내용

• 1-1교시

- 교수 소개
- 교과목 소개
- 강의 계획(강의계획서 참조)

• 1-2교시

- jamovi 란?
- 왜 jamovi를 배워야 하나

• 1-3교시

- jamovi 설치
- jamovi Tool 둘러보기

‘통계’란?

- 통계(statistics)

- 수집된 자료를 정리하고 그 내용을 특징지을 수 있는 수치를 산정하여 일정한 체계에 따라 그 결과값을 나타냄

- 결과 정보의 종류

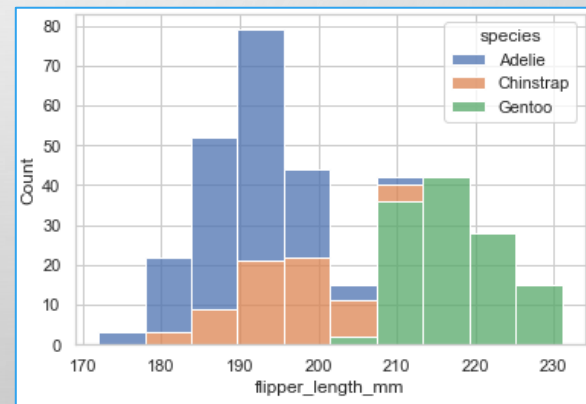
- 숫자

2, 3, 4, 5,
= {1, 2, 3, 4

- 문자

```
x a[5]={'a', 'b', 'c',  
x a[ ]={'a', 'b', 'c',  
x a[ ]={"abcd"};  
x a[5]={"abcd"};
```

- 그래프(시각화)



‘통계’란?

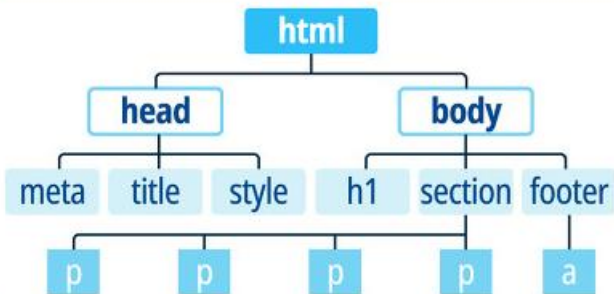
- 자료(data)
 - 정형 데이터 – 미리 정해 놓은 형식과 구조
 - 사용자가 쉽게 이해
 - 시스템에 쉬게 적용 가능
 - 이미 알고 있는 포맷형식 및 명확한 데이터 구조 표현
 - 반정형 데이터 – 데이터의 형식과 구조가 변경될 수 있음
 - 데이터의 구조 정보를 데이터와 함께 제공하는 파일 형식
 - 대표적인 반정형 데이터 구조
 - 하이퍼텍스트 마크업 언어(HTML)
 - 확장성 마크업 언어(XML)
 - 자원 기술 프레임워크(RDF)
 - 제이슨(JSON)
 - 비정형 데이터 – 구조가 없이 정형화되지 않은 데이터
 - 대표적인 비정형 구조
 - 동영상 파일
 - 오디오 파일
 - 사진
 - 보고서
 - 메일 내용

‘통계’란?

ID	Name	AGE	SEX
01	KIM	32	M
02	LEE	26	F
03	PARK	72	F
04	CHOI	15	M

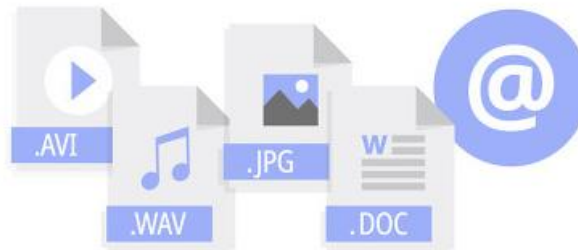
**structured
data**

정형데이터



**semi-
structured
data**

반정형데이터



**unstructured
data**

비정형데이터

통계분석을 하려면...

• 통계분석을 위한 툴

- SPSS, SAS, STATA, R, python, Excel
- SPSS : 사회과학 통계분석을 위해 만들어진 패키지
 - 특별히 깊이 있는 지식이 없어도 쉽게 사용할 수 있음
 - 표본이 크지 않아 설문조사, 실험연구 시 사용하기 적합
 - 하지만, 빅데이터를 처리하기에는 적합하지 않음
 - 다소 처리속도가 늦음
 - 수행했던 작업이 저장할 수 없음(처음부터 다시 시작)
- SAS : GUI 제공, 명령어 중심으로 사회과학분야 매우 많이 사용됨
 - University Edition 출시(무료버전)
 - 빅데이터 처리하는 데 성능이 뛰어남
 - 보건의료 분야에서 세계적으로 SAS 패키지를 통해 소통하고 있음



통계분석을 하려면...

• 통계분석을 위한 툴

- SPSS, SAS, STATA, R, python, Excel
- STATA : 계량경제학에 특화된 패키지로 사회과학보다 경제학에서 많이 사용됨
 - 처리속도가 매우 빠르고 명령어도 쉬움
 - 특히, 시계열이나 패널과 같은 동적 분석에 매우 탁월한 성능을 보임
 - 빅데이터 처리하는 데 매우 적합
 - 단, 계량경제학의 지식이 요구 됨
- R : 빅데이터를 활용한 머신러닝을 구축하는데 특화된 패키지
 - 무료버전, 오픈 라이브러리
 - 수많은 패키지 보유하고 있으며, 지속적으로 업데이트 되고 있음
 - 프로그램의 자유도가 높으므로 다양한 기능의 확장이 가능
 - 사용할 수 있으며 매우 활용성이 높음(AI관련)
 - 단, 프로그래밍언어를 배우는 시간이 필요함



통계분석을 하려면...

• 통계분석을 위한 툴

- SPSS, SAS, STATA, R, python, Excel
- python : 통계분석 뿐만 아니라 모든 분야에서 사용가능(범용성).
 - 유연성과 호환성이 매우 뛰어남
 - 다양한 작업을 할 수 있음
 - AI 구현에 있어 핵심적 요소
 - 다른 컴퓨터 언어에 비해 배우기 수월함
- Excel : 기초적인 통계분석 가능
 - 추가 기능을 통해 메뉴를 추가할 수 있음
 - 자료 구조가 잘 정리되면 간단하게 결과를 얻을 수 있음
 - 많은 기능을 제공하고 있음



‘jamovi’ 란?

• ‘Jamovi’란?

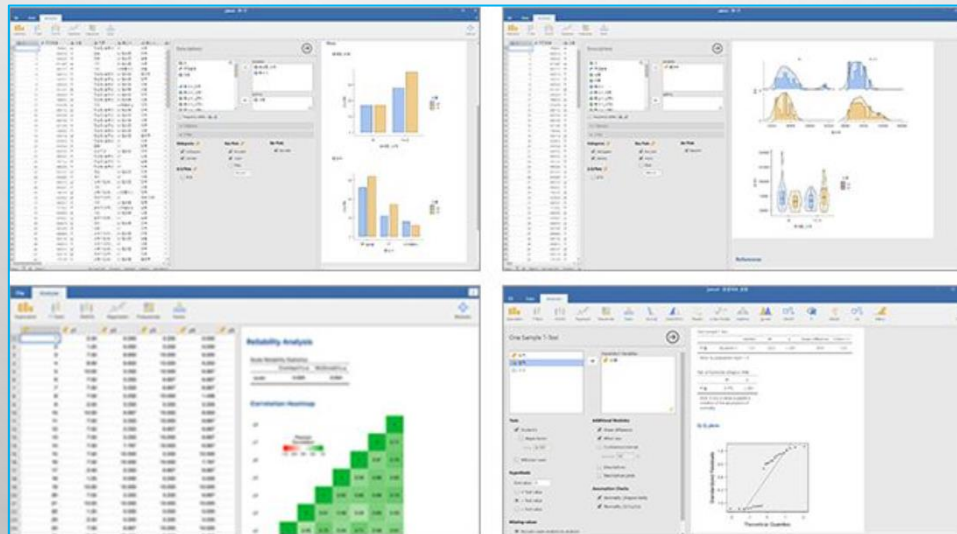
- 통계 분석기능과 함께 통계 모듈(패키지) 개발 환경을 제공함
- R 프로그래밍 언어로 작성되었음
- R 패키지 개발 경험이 있다면..
 - Jmvtools를 사용해서 직접 통계 모듈을 개발하여 jamovi 통계 프로그램에 설치하고 실행할 수 있음
 - 통계모듈 개발 시, Jamovi 홈페이지에 있는 개발 메뉴얼 참조할 것



‘jamovi’ 란?

• 왜 ‘Jamovi’를 배울까?

- SPSS에 있는 기능을 모두 가지고 있음
- Excel 처럼 데이터 불러와서 클릭만 하면 분석 됨
- 고급 기능 필요시 R 프로그램과 연동됨으로 편리
- 사회과학 논문 결과 보고 양식의 기준인 APA 스타일로 제공
 - 사용자가 분석 결과를 쉽게 이해할 수 있음
- 편집과정 없이 논문에 분석결과를 제시할 수 있는 환경 제공
- GUI 방식으로 초보자도 쉽게 사용할 수 있음
- Jamovi 프로그램은 무료 오픈 소스(python 과 같음)

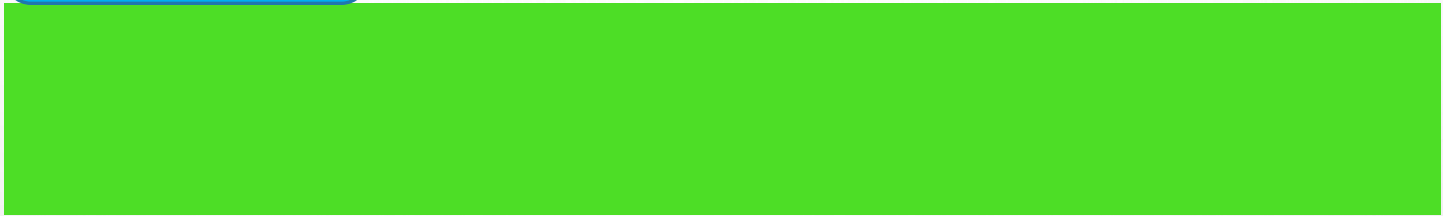


빅데이터 분석 기법의 종류

기초이론		도구	데이터 관리	시각화	기술통계
빅데이터 개요 ① 활용 ② 전망		① R ② Rstudio ③ JAVA	① 데이터 입력 ② 데이터 출력 ③ 데이터 처리	① 기본 ② 응용	① 빈도 ② 교차
기초이론		T-test		ANOVA	상관분석
빅데이터 수집/저장	통계분석 이해	① 일표본 ② 독립표본 ③ 대응표본		① 일원배치 ② 이원배치	① 피어슨 ② 스피어,
예측분석. 1		예측분석. 2		기초이론	예측분석. 3
선형회귀분석 ① 다중회귀분석 ② 매개효과분석		① 구조방정식모형분석 ② 조절효과분석		데이터 마이닝 이해	① 이항 로지스틱 ② 다항 로지스틱
분류분석. 1		분류분석. 2	분류분석. 3		분류분석. 4
판별분석 ① 선형판별분석 ② 이차선형판별분석		① 의사결정 트리(DT)	① 랜덤포레스트 (RF)		① 서포트벡터머신 (SVM)
기초이론	클러스터링	클러스터링	연관성분석		시계열분석
텍스트 마이닝 이해	① 계층적 분석	② 비계층적 분석	① 연관규칙, ② 순차적 패턴분석		① 정상시계열(ARIMA) ② 비정상시계열 (ARIMA)
토픽분석	감성분석	DEA분석	기초이론	DNN	
① 워드 클라우드 분석	① 감성 분석	① 효율성 분석	딥러닝 이해	① 심층신경망분석 (DNN)	
					연구 모델링

<1-2차시, 자기주도적 학습정리>

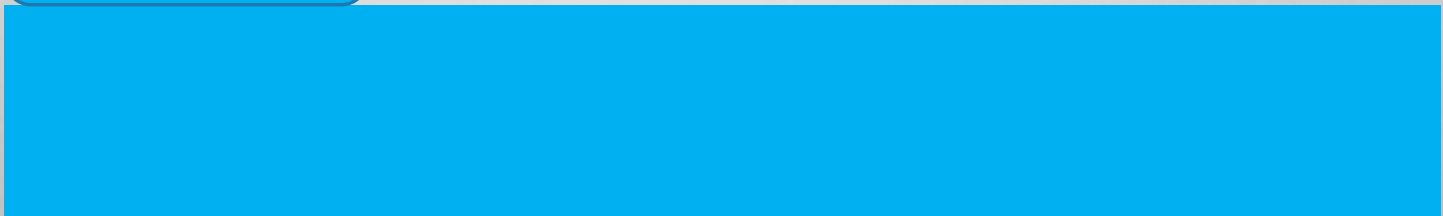
PC! 구성요소



PC! 동작순서



PC! data 처리



대면 강의 내용(1주차)

- 1주차 강의내용

- 1-3교시

- jamovi 설치
 - jamovi Tool 둘러보기

‘코로나19’로 대면-비대면 강의 진행

수고하셨습니다