

2023학년도 2학기 언어데이터과학

제6강 데이터과학 (2)

박수민

서울대학교 인문대학 언어학과

2023년 9월 20일 수요일

지난 시간에 배운 것

- 1 데이터의 정의, 네 가지 종류, 언어데이터의 범위
- 2 기초통계량(평균, 중앙값, 편차, 표준편차, 공분산, 상관관계)

오늘의 목표

- 1 지난 시간에 배운 기초통계량을 Python으로 구현할 수 있다.
- 2 numpy에서 배열(array) 자료형의 객체를 만들고 연산할 수 있다.

지금 할 일

- 1 VS Code에서 github LDS2023 레포지토리를 연동한다.
- 2 아래 URL의 notes/ 폴더에서 오늘 날짜에 해당하는 파일들을 다운로드한다.
 - <https://github.com/suparklingmin/LingDataSci2023>

참고: 기존 컴퓨터에서 LDS2023 레포지토리에 접근하는 방법

- 1 Activity bar > [Source Control] > [Open Folder] 클릭.
- 2 지난번에 사용한 폴더 선택.

참고: 새로운 컴퓨터에서 LDS2023 레포지토리를 복제해 오는 방법

- 1 Activity bar > [Source Control] > [Clone Repository] 클릭.
- 2 Clone from GitHub 클릭.
- 3 username/LDS2023 입력.

마지막으로 할 일

- 1 노트 파일을 username/LDS2023/notes 폴더에 저장한다.
- 2 Commit & Push를 완료한다.

다음 주에 배울 것

- 1 pandas 모듈과 탐색적 자료 분석
- 2 seaborn 모듈과 데이터 시각화

[숙제03] numpy 연습

- 1 numpy 라이브러리를 활용하여 기술통계량 구현하기.
- 2 연습문제 풀기.