|  |  |
| --- | --- |
| 교육 제목 | AI 프로젝트 기반 S/W 전문가 양성과정 |
| 교육 일시 | 21.10.27 |
| 교육 장소 | 대면(영우글로벌러닝 강의실) |
| **교육 내용** | |
| 오전 | 1.작일 수업 리뷰 <Git 복습>  # file 생성 시 업로드 절차  # git 페이지에서 수정되어 있는 사항 로컬 적용  # 서버와 로컬의 충돌. git과 로컬에 특정 구간 이후 상이한 부분이 발생할 때  시퀀스 중 기존 알고 있던 내용에서 수정  git stauts  git add  git commit  git push -> Error가 발생 하면 서버와의 동기화 절차 시작  git pull  ->Automatic merge failed. 머지 중에 틀린 부분을 모두 마지막 부분에  붙여 놓음. 사람이 직접 머지 해야함  ->Head : local에서 수정 한 부분  ->Hesh : 서버에서 수정된 부분  -> local에서 수정하여 업로드  git add  git commit  git push    # branch 사용  # branch 생성  # 브런치 확인  # 브런치 이동  #새로 만든 브런치에 파일 생성 및 브런치 적용    2. 통계  #연속형 자료, 범주형 자료  #표본집단, 모집단  #범주형 자료를 표현하기 좋은 그래프  #연속형 자료  돗수분포표->범위를 정하고 평균값을 계급 값으로 지정. 도수를 표로 정리  히스토그램->분포의 비율을 시각화 한 그래프  상자수염그림, 바이올린그림 -> 데이터 시각화에 특화  #그래프를 통한 자료요약의 장단점  장점: 자료를 한 눈에 알아볼 수 있음  단점: 그림의 모양이 작성자의 주관적 판단에 따라 달라 질 수 있음 |
| 오후 | #수치를 통한 연속형 자료의 요약  표본평균  중심을 나타내는 측도 중에서 가장 많이 사용되는 방법  자료의 무게 중심을 나타냄  자료의 이상치 (Outlier)에 영향을 많이 받음 (Trimmed mean)  중앙값  전체 관측값을 크기 순으로 나열한 했을 때 중앙에 위치한 값  자료의 이상치 (Outlier)에 영향을 적게 받음  # 표본평균, 중앙값, 최빈값의 비교  표본평균은 이해하기 쉽고 이론적 전개가 용이  표본평균은 전체 관측치를 반영하지만 이상치 (Outlier)에 영향을 받음  중앙값 중앙 부분의 관측치에 영향을 받고 이상치 (Outlier)에 영향을 받지 않음  이상치 들이 있는 경우 표본평균과 중앙값을 적절히 사용    # 퍼짐의 정도  자료가 중심으로 부터 얼마나 퍼져있는 지를 표현 수치  표준편차 sd = √s2 = S  범위(Range): 최댓값 - 최솟값  사분위수범위: 3사분위수 - 1사분위수 -> 데이터의 퍼짐 정도  # 상관분석 (Correlation)  두 연속형 변수간에 선형적 연관관계가 있는 지를 분석하는 통계적 방법  모집단 상관계수는 ρ, 표본집단의 상관계수는 r로 표기  # 피어슨 상관계수 (Pearson’s r)  두 연속형 변수가 정규분포를 따르는 경우에 사용함  rp = Cov(x,y) / ( √(Var(x)) \* √(Var(y)) )  r값의 범위는 -1~+1 까지 분포하며 절대값이 1에 가까울 수록 높은 상관성  절대값이 0.7 ~ 0.9 : 높은 상관 관계  절대값이 0.4 ~ 0.7 : 중등도 상관 관계  #스피어만 상관계수 (Spearman’s r)  순서통계량, 두 연속형 변수가 정규분포를 따르지 않는 경우에 사용함  두 변수 간에 상관관계를 나타내는 비모수적 방법  순서형 자료에서도 적용이 가능, 비선형적 연관성도 판단할 수 있음  r 값의범위 -1~+1 분포하며 절대값이 1에 가까울 수록 높은 상관성 |