|  |  |
| --- | --- |
| 교육 제목 | 개발기초 역량교육\_디지털 트랜스포메이션\_2 |
| 교육 일시 | 2021.09.15 |
| 교육 장소 | 영우글로벌러닝 |
| **교육 내용** | |
| 오전 | 1. 어제 배운 git 명령어 복습   1-1. repository 생성  1-2 클론, 파일 생성  1-3 Branch 로컬에서 생성, 클라우드 master와 충돌시 해결   1. 통계   2-1. 자료(정형 data)의 종류. 연속/이산(범주).  연속: interval(등간), ratio(비율) 이산(범주): nominal(명목), ordinal(순서)  2-2. population(모집단):이해하고자 하는 집단 전체 sample space(표본집단): 모집단에서 특정 목적에 따라 추출된 집단  2-3. 시각화 표현: 범주형(이산) contingency table(kxc 분할표), barplot(막대그래프), pie chart(파이 차트)  연속형: frequency table(돗수분포표), histogram(히스토그램), boxplot(상자수염 그림), Violin plot(바이올린 그림)  Violin plot은 측정값이 어떤 분포를 하는지 알 수 있어 평균/중앙값 중 뭐가 더 유의미하게 쓰일지 알 수 있다.  2-4 시각화 표현의 장단점: 장점은 자료를 한눈에 파악하기 좋고, 단점으로는 작성자가 의도적으로 눈금 후려치기를 시도하기도 한다.   1. 수치를 통한 연속적 자료의 요약   3-1. 표본평균(sample mean) mean=x\_i\*p\_i ; x\_i=표본, p\_i=확률. Weight이 있어서 outlier에 영향을 크게 받음  3-2 중앙값(median): f(a)=sum|x\_i-a| f(a)를 최소화하는 값. 중앙값. 이산 데이터에 대해, 홀수면 m=(n+1)/2번째 데이터 짝수면 n/2번째와 (n+1)/2번째 데이터의 평균 weight을 없애서 outlier에 영향을 적게 받음  3-3 최빈값(mode): 가장 많이 나온 값   1. 퍼짐의 정도   4-1. 분산은 중심으로부터 얼마나 퍼져있는지 나타냄 (평균값으로부터 떨어진 정도를 나타내기 위해, 절대값을 활용하기 위해 제곱)  4-2. 표준편차는 루트를 활용 단위 길이를 맞춰줌  4-3. Range(범위):x\_max-x\_min IQR(사분위수범위):Q\_3-Q\_1 |
| 오후 | ※ 표본평균은 표준편차와, 중앙값은 IQR과 짝지어 쓰인다.     1. 두 변수간의 자료 요약   5-1 상관분석: 연관성만 봄. 인과성이 아님에 유의. 인과성은 회귀분석(regression)  5-2. 피어슨 상관계수: 정규분포데이터, 실제 값 사용  5-3. 스피어만 상관계수: 비정규분포 데이터, rank 사용   1. 확률   6-1 표본공간과 그 원소 근원사건. 특정 사건이 주어지면 표본공간의 부분집합으로 사건이 특정됨  6-2 확률은 n이 무한대고 실수 하나로 특정. 실질적으로는 n에따라 분산식이 주어짐.  6-3 조건부확률. 사건 A와 B가 독립일때는 교집합 확률이 각자 확률의 곱으로 나타남.  6-4 독립인지 아닌지 판별할 때 실질적으로는 독립성검증(카이제곱 분포)를 이용한다. |