|  |  |
| --- | --- |
| 교육 제목 | **기초 통계학** |
| 교육 일시 | 2021-09-15 |
| 교육 장소 | 집 |
| **교육 내용** | |
| 오전 | 깃허브 자료 업로드/다운로드  파일로 이동  git clone 주소  git add 파일  git commit -m "내용"  git push  git fetch  git pull    충돌시  git pull  git push  내용수정후  git pull  git add 파일  git commit -m "내용"  git push  git checkout -b branch\_new 생성  git branch branch 현재 사용자 확인  로컬 파일 수정  git add 파일  git commit -m "수정내용"  git checkout master 마스터로 변경  git push origin branch\_new 브렌치 변경사항 확인  정형 데이터  기초 통계학  자료 종류  연속형 자료  1. 등간척도 : 절대값 0 존재 x , 예)온도 -> 물리적으로 곱셈법칙 적용 안됨 물리적 의미로 0은 존재 가 없는것 온도 1 과 2는 수치상 2배 가 아님  2. 비율척도 : 절대값 0 존재 예)키, 몸무게, 혈압  자료 -> 보통 정형 자료 (정형 : 정리된 자료)  나누는 이유 중요 하지않음, 방식은 다르지않음  하지만 의미적 차이는 있음    범주형 자료  1. 명도척도 : 속성을 순위없이 분류 예) 남여, 국어영어수학  2. 순서척도 : 속성간 순위 존재 예) 상중하, 좋음/보통/나쁨 |
| 오후 | 비정형  그림, 글, 단어, 비전vision -> NLP 등  -------------------------------------  자료 축약  모집단 : 집단 전체  표본집단 : 추출된 집단  -------------------------------------  연속형 자료형  돗수 : 범주에 속하는 값의 갯수와 그 범주  상대도수 : 도수를 전체 자료수로 나눈 값  도수분포표 : 범주와 도수, 상대도수를 나타낸 표  종류  히스토그램  상자수염 그림  바이올린 그림  ---------------------------------------  그래프 장단점  시각화되어 이해하기 쉬움, 하지만 작성자의 주관적이기때문에 판단 달라질수도 있음  연속형 자료 요약  중심위치 측도  표본평균 : 가장많이 사용,  중앙값 : 크기순으로 중앙에 위치한 값 ( 홀수 (n +1)/2 번쨰 값, 짝수 (n/2)+(n+1)/2)  최빈값 : 관측값 중 가장 자주나온 값  표본평균 예  #A tibble : 5 \* 2  value n  <dbl> <int>  1 3 1  2 5 2  3 8 1  4 9 1  5 10 1  = 3 + 5 + 5 +8 + 9 10 / 6 = 6.67 표본평균(mean)  ((6/2)+(6+1)/2) = 6.5 중앙값(median)    퍼짐 정도  분산, 표준편차  상관분석  두개이상 연속형 변수간 연관관계 표현/분석  그림으로 -> 산점도  피어슨 상관계수  두연속형 변수가 정규분포를 따르는 경우 사용  테스트방법은 따로있음 대표적으로 히스트 함수  스피어만 상관계수  두 변수 간 상관관계 비모수적 방법 표현  ------------------------------------------  확률 : 동일 실험 무한히 반복한 결과  사건의 확률  표본공간 : 한 실험에서 나올수 있는 결과 모음  근원사건 : 표본공간 구성하는 결과  사건 : 표본공간 부분집합 근원 사건들의 집합  확률 계산 방법  1. 여사건의 법칙  2. 합사건의 법칙  3. 합사건의 여사건  4. 조건 부확률  5. 두사건 A,B가 독립인경우  확률 분포  확률 변수 |