|  |  |
| --- | --- |
| 교육 제목 | **GIT 복습 및 통계 기초** |
| 교육 일시 | 2021년 10월 27일 |
| 교육 장소 | 영우글로벌러닝 2층 |
| **교육 내용** | |
| 오전 | 1. git 사용법 복습 2. Basic statistics 기초통계학 3. 자료의 종류 4. 범주형 자료, classification 분류    1. 명목척도 (Nominal): 학생 번호    2. 순서척도 (Ordinal): 성적 5. 연속형 자료, regression, 회귀    1. 등간척도 (Interval): 온도    2. 비율척도 (Ratio): 키, 몸무게, 혈압 6. 자료의 축약 7. 모집단과 표본집단    1. 모집단: 이해하고자 하는 집단전체, Bigdata 대상, P-value 나 확률 등의 용어를 거의 사용하지 않게 됨    2. 표본집단 : 모집단에서 추출된 집단, 8. 범주형 자료의 요약    1. RxC 분할표 (Contingency table)    2. 막대그래프 (Barplot)    3. 파이 차트 (Pie chart) 9. 연속형 자료의 요약    1. 돗수분포표 (Frequency table)    2. 히스토그램 (Histogram) : 연속성 자료!    3. 상자수염 그림 (Boxplot)    4. 바이올린 그림 (Violin plot) 10. 그래프를 통한 자료의 요약의 장단점     1. 장점: 자료를 한 눈에 알아볼 수 있음     2. 단점: 그림의 모양이 작성자의 주관적 판단에 따라 달라질 수 있음 조작의 가능성이 농후.  따라서 수치 데이터를 기초 자료로 함께 사용(신뢰도) 11. 수치를 통한 연속형 자료의 요약     1. 중심위치의 측도     2. 표본평균 (Sample Mean)        1. 중심을 나타내는 측도 중에서 가장 많이 사용되는 방법        2. 자료의 이상치 (Outlier)에 영향을 많이 받음 (Trimmed mean)     3. 중앙값 (Median)        1. 전체 관측값을 크기 순으로 나열한 했을 때 중앙에 위치한 값        2. 자료의 이상치 (Outlier)에 영향을 적게 받음     4. 최빈값 (Mode)        1. 관측값 가운데 가장 자주 나온 값 |

|  |  |
| --- | --- |
| 오후 | 1. 표본평균, 중앙값, 최빈값의 비교    1. 표본평균은 이해하기 쉽고 이론적 전개가 용이    2. 표본평균은 전체 관측치를 반영하지만 이상치 (Outlier)에 영향을 받음    3. 중앙값 중앙 부분의 관측치에 영향을 받고 이상치 (Outlier)에 영향을 받지 않음    4. 이상치 들이 있는 경우 표본평균과 중앙값을 적절히 사용 2. 퍼짐의 정도: 분산과 표준편차 (Variance and Standard deviation, SD)    1. 자료가 중심으로 부터 얼마나 퍼져있는 지를 표현 수치    2. 사분위수범위(InterQuartile Range, IQR): 3사분위수 - 1사분위수 3. 표본편차, 범위, 사분위수범위 비교    1. 표준편차는 표본평균과 같은 이론적 배경    2. 사분위수 범위는 중앙값과 같은 이론적 배경    3. 중심측도로서 표본평균를 사용할 경우 표준편차를 사용    4. 중심측도로서 중앙값을 사용할 경우 사분위수 범위를 사용 4. 상관분석 (Correlation)    1. 두 연속형 변수간에 선형적 연관관계가 있는 지를 분석하는 통계적 방법    2. 두 연속형 변수간의 연관된 정도를 나타내는 척도이며 인과관계를 설명하는 것은 아님 - 상관계수(Correlation coefficient)는 선형적 관계의 강도를 나타냄    3. 모집단 상관계수는 ρ, 표본집단의 상관계수는 r로 표기 5. 피어슨 상관계수 (Pearson’s r)    1. 두 연속형 변수가 정규분포를 따르는 경우에 사용함    2. r 값이 + 이면 양의 상관관계, - 값이면 음의 상관 관계를 의미    3. r 값의 범위는 -1 에서 +1 까지 분포하며 절대값이 1에 가까울 수록 높은 상관성을 의미    4. r 값의 해석 (절대적 기준은 아님)    5. 절대값이 0.7 ~ 0.9 : 높은 상관 관계    6. 절대값이 0.4 ~ 0.7 : 중등도 상관 관계 |