|  |  |
| --- | --- |
| 교육 제목 | **기초 통계학** |
| 교육 일시 | 2021-09-15 |
| 교육 장소 | 비대면 |
| **교육 내용** | |
| 오전 | **자료의 종류**  -연속형 자료  등간 척도 (Interval): 절대값 0이 존재하지 않음, 곱셈법칙 적용이 안됨 예)온도  비율척도 (Ratio): 절대값 0이 존재 예) 키, 몸무게, 혈압  -범주형 자료  명목척도 (Nominal): 속성을 분류하는 척도, 속성간 순위가 없음, 예) 남/녀, 국어/영어/수학  순서척도 (Ordinal): 속성 간에 순위개념이 존재하는 척도, 예) 상/중/하, 좋음/보통/나쁨  **-모집단과 표본집단**  모집단: 이해하고자 하는 집단전체  표본집단 : 모집단에서 추출된 집단  -범주형 자료의 요약  k x c 분할표,막대그래프,파이 차트  -속형 자료의 요약  돗수분포표 (Frequency table)  ==범주에 속하는 관측값의 갯수를 그 범주의 돗수라고 함  이 돗수를 전체 자료의 개수로 나눈 값을 상대돗수라고 함  각 범주에 대응되는 돗수와 상대돗수를 나타낸 표===  히스토그램,상자수염 그림 ,바이올린 그림  -그래프를 통한 자료의 요약의 장단점  장점: 자료를 한 눈에 알아볼 수 있음  단점: 그림의 모양이 작성자의 주관적 판단에 따라 달라질 수 있음  **-표본평균 (Sample Mean)**  중심을 나타내는 측도 중에서 가장 많이 사용되는 방법  자료의 무게 중심을 나타냄  자료의 이상치 (Outlier)에 영향을 많이 받음 (Trimmed mean)  **-중앙값 (Median)**  전체 관측값 을 크기 순으로 나열한 했을 때 중앙에 위치한 값  Median (m): P(X≤m)=∫m−∞f(x)dx=12  **-최빈값 (Mode)**  관측값 가운데 가장 자주 나온 값  **-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**  분산과 표준편차 (Variance and Standard deviation, SD)  자료가 중심으로 부 터 얼마나 퍼져 있는 지를 표현 수치  범위(Range): 최댓값 - 최솟값  사분위수범위 (InterQuartile Range, IQR): 3사분위수 - 1사분위수  **-표본편차, 범위, 사분위수범위 비교**  표준편차는 표본평균과 같은 이론적 배경  사분위수 범위는 중앙값과 같은 이론적 배경  중심측도로서 표본평균를 사용할 경우 표준편차를 사용  중심측도로서 중앙값을 사용할 경우 사분위수 범위를 사용  -상관분석 (Correlation)  두 연속형 변수간에 선형적 연관관계가 있는 지를 분석하는 통계적 방법  두 연속형 변수간의 연관된 정도를 나타내는 척도이며 인과관계를 설명하는 것은 아님 - 상관계수(Correlation coefficient)는 선형적 관계의 강도를 나타냄  모집단 상관계수는 ρ, 표본집단의 상관계수는 r로 표기  그림을 통한 상관관계 요약: 산점도 |
| 오후 | -**피어슨 상관계수**  연속형 변수가 정규분포를 따를 때  원래데이터를 직접가지고  두 연속형 변수가 정규분포를 따르는 경우에 사용함  r 값이 + 이면 양의 상관관계, - 값이면 음의 상관 관계를 의미  r 값의 범위는 -1 에서 +1 까지 분포하며 절대값이 1에 가까울 수록 높은 상관성을 의미  r 값의 해석 (절대적 기준은 아님)  절대값이 0.7 ~ 0.9 : 높은 상관 관계  절대값이 0.4 ~ 0.7 : 중등도 상관 관계  -**스피어만 상관계수**  랭크 값 으 로 계산  두 변수 간에 상관관계를 나타내는 비모수적 방법  두 연속형 변수가 정규분포를 따르지 않는 경우에 사용함  순서형 자료에서도 적용이 가능  비선형적 연관성도 판단할 수 있음  r 값의 범위는 -1 에서 +1 까지 분포하며 절대값이 1에 가까울 수록 높은 상관성을 의미  확률(동일한 실험을 무한히 반복했을 때 나타나는 사건의 상대 돗수의 비)  확률의 법칙 통계확률 은0-1 확률공간만 다룸  모든 사건 A 에 대하여 0≤P(A)≤1  P(A)=∑ωi∈AP(ωi)  P(Ω)=∑ωi∈ΩP(ωi)=1  확률의 계산  여사건의 법칙: P(Ac)=1−P(A)  합사건의 법칙: P(A∪B)=P(A)+P(B)−P(A∩B)  합사건의 여사건: P((A∪B)c)=P(Ac∩Bc)  조건부확률: P(A|B)=P(A∩B)P(B), P(B)>0  두 사건 A,B가 독립인 경우: P(A∩B)=P(A)P(B)    테이터 의 n수에 따른 결정하는 카이 제곱 분포가 있다  근사치는 카이 제곱 테스트를 해봐야 정확히 알 수 있다 |