

# 확률 및 통계 1차 과제물

소프트웨어학부 2021041017 김규현

[1] 어느 도시에서 실업률을 알아보려고 그 도시의 전체 노동 인구 10만 중 500명을 추출하여 이들의 실업률을 조사함으로써 그 도시의 전체 실업률을 알아보려고 한다. 500명의 실업률을 조사한 결과 실업률은 8.0%였다. 여기서 모집단, 모수, 표본, 통계량을 정의하라.

모집단(population) : 노동 인구 10만 명

모수(parameter) : 도시 전체 노동 인구의 실업률

표본(sample) : 500명의 조사된 노동자

통계량(Statistic) : 실업률 8.0%

[2] 다음 자료를 연속형, 이산형, 순서척도, 명목척도, 등간척도로 분류하고 이유를 써라.

(1) 주유할 때 선택하는 휘발유의 특정상호 : 특정 상호는 단순히 구분을 위한 이름을 뿐이므로, 순서가 없는 범주형 데이터인 명목척도이다.

(2) 서울에서 지방으로 출장을 갈 때 사용하는 교통수단 : 교통수단은 순서가 없이 이름으로만 구분되는 범주형 데이터로, 순서가 없으므로 명목척도이다

(3) 일일 최고기온 : 기온은 연속적으로 측정되며, 측정치 사이의 거리가 일정하므로, 연속형 데이터, 등간척도이다.

(4) A대학교의 특정학과 1학기 등록금 : 등록금은 연속적인 값으로 측정되고, 절대적 원점이 존재하며 비율 계산이 가능하므로 연속형 자료, 비율 척도이다.

(5) 신체검사 때 측정한 혈압 : 혈압은 연속적으로 측정되는 수치고, 절대적 원점을 가질 수 있으며 비율을 계산할 수 있으므로, 연속형 데이터, 비율 척도이다.

(6) 특정회사의 커피 향을 5등급으로 결정할 때의 결과 값 : 5등급 평가는 범주형 데이터로, 등급 간 순서가 있으므로 순서척도이다.

(7) 완성된 제품의 품질 검사 결과 : 품질 검사 결과는 범주형 데이터로, 각 case마다 명목, 순서 척도로 나눌 수 있다. 예를 들어, '불량/양호'는 순서가 없으므로 명목척도, '1~5등급'으로 나눈다면 순서척도가 될 수 있다.

(8) 국회의원의 FTA 보호무역에 대한 선호도 : 선호도를 측정하는데 '매우불만=1, 불만=2 ...' 와 같이 측정치 사이의 거리를 일정한 간격으로 표시하면 연속형 데이터인 등간척도가 될 수 있고, '매우 찬성, 찬성...'와 같이 표시하면 범주형 데이터의 순서척도가 될 수 있다.

(9) 지난 밤 발생한 사건사고의 수 : 사건사고의 수는 이산형 데이터이고, 절대적 원점이 존재하며, 사건사고의 수를 비율로 표시할 수 있으므로 비율척도이다.

(10) 기초통계학에서 얻은 학점(A, B, C, D, F) : 학점은 순서가 있는 범주형 데이터로, 순서척도이다

[3] 종로산업(주)의 품질관리부는 오븐을 생산하는 세 개의 조립라인을 항상 감시한다. 오븐은 4분 안에 화씨 40도까지 가열한 후 끄도록 설계되어 있다. 그런데 차단시설의 잘못으로 4분 안에 240도에 이르지도 못하거나 반대로 이를 초과하는 경우도 발생한다. 생산라인으로부터 온도에 관한 많은 표본을 추출한 결과 다음과 같은 자료를 얻었다.

통계 측정치	라인1	라인2	라인3
평균	238.1	240.0	242.9
중앙값	240.0	240.0	240.0
최빈치	241.5	240.0	239.1
표준편차	3.0	0.4	3.9
사분위수 범위	2.0	0.2	3.4

**(1) 어떤 라인이 종모양의 분포에 가까울까 ?**

종모양의 분포는 정규분포를 의미하며, 다음과 같은 특징을 가진다.

평균( $\mu$ ), 중앙값(median), 최빈치(mode)가 비슷할수록 종모양 분포에 가깝다.

라인 2의 경우: 평균 = 240.0, 중앙값 = 240.0, 최빈치 = 240.0,

평균, 중앙값, 최빈치가 모두 동일하므로 라인 2가 정규분포(종모양의 분포)에 가장 가깝다.

**(2) 온도에 있어 심한 변동을 나타내는 라인은?**

심한 변동을 나타내는 라인을 찾으려면 변동계수가 가장 큰 라인을 찾으면 된다.

변동계수(CV) = 표준편차 / 평균

라인 1의 경우 :  $3.0 / 238.1 \approx 0.0126$

라인 2의 경우 :  $0.4 / 240.0 \approx 0.0017$

라인 3의 경우 :  $3.9 / 242.9 \approx 0.0161$

∴ 라인 3의 변동계수가 가장 크므로, 심한 변동을 나타내는 라인은 라인3이다.

**(3) 온도 분포가 오른쪽 꼬리 모양을 갖는 라인은?**

자료의 분포 모양에서 오른쪽 꼬리 모양을 갖는 것은 평균 < 중앙값의 경우이다.

라인 1~3중 '라인 1'만이 평균 238.1 < 중앙값 240.0 이므로, 라인 1이 오른쪽 꼬리 모양을 갖는다.

**(4) 라인 2의 1사분위수, 3사분위수를 구하라.**

라인 2의 1사분위수(Q1) : 하위 25% 값, 3사분위수(Q3) : 상위 25% 값

사분위수 범위(IQR) =  $Q3 - Q1 = 0.2$

$Q1 = Q2 - (IQR / 2) = 240.0 - (0.2 / 2) = 240.0 - 0.1 = 239.9$

$Q3 = Q2 + (IQR / 2) = 240.0 + (0.2 / 2) = 240.0 + 0.1 = 240.1$

(5) 라인 1에서 비대칭도는 얼마인가?

$$\text{비대칭도} = S_k = \frac{3(\bar{x} - M_d)}{s}$$

$S_k = 0$ (대칭),  $S_k < 0$ (왼쪽 비대칭),  $S_k > 0$ (오른쪽 비대칭)

비대칭도 공식(출처 : 강의노트)

비대칭도 = 3(평균 - 중앙값) / 표준편차

라인 1의 평균 : 238.1, 중앙값 : 240.0, 표준편차 : 3.0

라인 1의 비대칭도 =  $3(238.1 - 240.0) / 3.0 \approx -1.9$

추가로 알 수 있는 사실 : 라인 1의 비대칭도가 음수이므로 왼쪽으로 비대칭이라는 점을 알 수 있다.

~~[4] 3장 강의노트 20페이지의 베이즈정리 3문제를 풀어볼 것(풀이과정 상세히)~~

**진도 문제로 4번은 생략.**