확률 및 통계 1차 과제물

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

답안)

- 모집단 : 어떤 정보를 얻기 위해 연구대상으로 관심을 두고 있는 전체 집단
  → 도시 전체의 노동 인구 10만 명
- 모수 : 모집단 특성을 나타내는 미지의 값 → 모비율~도시 전체 실업률
- 표본 : 모집단 특성에 관한 정보를 얻기 위해 모집단으로부터 추출한 값 또는 측정한 값들에 대한 집합 --> 노농인구 10만명 중 추출 된 500명
- 통계량 : 표본의 특성을 수치로 나타낸 값 --> 표본비율~ 실업률 8.0%

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

(1) 주유할 때 선택하는 휘발유의 특정상호

(2) 서울에서 지방으로 출장을 갈 때 사용하는 교통수단

(3) 일일 최고기온

(4) A대학교의 특정학과의 1학기 등록금

(5) 신체검사 때 측정한 혈압

(6) 특정회사의 커피 향을 5등급으로 결정할 때의 결과 값

(7) 완성된 제품의 품질 검사 결과

(8) 국회의원의 FTA 보호무역에 대한 선호도

(9) 지난 밤 발생한 사건사고의 수

(10) 기초통계학에서 얻은 학점(A, B, C, D, F)

답안)

(1) 주유할 때 선택하는 휘발유의 특정상호
→ 명목척도 : 특정 상호의 분류목적의 자료.

(2) 서울에서 지방으로 출장을 갈 때 사용하는 교통수단
→ 명목척도 : 교통수단의 분류목적의 자료.

(3) 일일 최고기온
→ 연속형 : 특정 온도범위(구간)에서 무한한 실수값으로 관측가능.
ex) 섭씨 온도(연속형이지만 비율척도는 아니다- 구간척도)

(4) A대학교의 특정학과의 1학기 등록금

→ 연속형 : 특정 범위(예:250만~800만)에서 무한한 실수 값으로 관측가능. ~ 비율척도 가능.

(5) 신체검사 때 측정한 혈압

→ 연속형 : 특정 범위(혈압)에서 무한한 실수 값으로 관측가능. ~ 비율척도 가능

(6) 특정회사의 커피 향을 5등급으로 결정할 때의 결과 값

→ 순서척도 또는 등간척도 : 순서척도

(7) 완성된 제품의 품질 검사 결과

→ 순위척도 : 미흡(1), 보통(2), 우수(3)~품질 상태에 대한 서열이 가능. ~ 순서척도

(8) 국회의원의 FTA 보호무역에 대한 선호도

→ 순서 또는 등간척도 : 순서척도이며 등 간격(선호도 점수 1~5점)자료이면 등간척도도 가능.

(9) 지난 밤 발생한 사건사고의 수

→ 이산형 : 셀수 있는 형태의 자료. ~ 양적자료의 이산형 ~ 비율척도 가능

(10) 기초통계학에서 얻은 학점(A, B, C, D, F)

→ 순서척도 / 90이상(A), 80이상(B), 70이상(C), 60이상(D), 50미만(F).

[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 ?

(2) 온도에 있어 심한 변동을 나타내는 라인은 ?

(3) 온도분포가 오른쪽 꼬리모양을 갖는 라인은 ?

(4) 라인 2의 1사분위수, 3사분위수를 구하라.

(5) 라인 1에서 비대칭도는 얼마인가 ?

답안)

① 라인2($\overline{X} = Md = Mo$) ② 라인3(변동계수=0.016) ③ 라인3($Sk = 2.23$)

④ 239.9, 240.1 ⑤ $Sk = -1.9$ 왼쪽꼬리분포

확률 및 통계 2차 과제물

예1) (사전확률 정보) 발병률이 0.05인 A질병이 있다고 하자. 실제로 A병에 걸린 사람이 검사하면 0.90의 비율로 양성판정, 병에 걸리지 않은 사람이 검사하면 0.10의 비율로 양성판정으로 나온다고 한다. 그렇다면 이 질병 검사에서 양성판정이 나왔을 때, (사후확률 추정) 검사 대상자가 진짜 질병에 걸렸을 확률은 ?

1. 사전확률 : $P(A)=0.05$, $P(A^c)=0.95$

2. B: 양성

$P(B|A)=0.9$, $P(B|A^c)=0.1$

3. 사후확률 : $P(A|B) = \frac{P(A\cap B)}{P(B)} = \frac{P(B|A)\cdot P(A)}{P(B)}$

$$= \frac{0.9\times 0.05}{0.05\times 0.9 + 0.95\times 0.1}$$

$$= \frac{0.045}{0.045+0.095} = \frac{0.045}{0.14} = 0.32$$

예2) 실험자 학생이 거짓말을 하는지 알아보기 위해 거짓말 탐지기를 사용하기로 했다. (사전확률 정보) 학생이 거짓말을 할 확률이 0.25이고, 진실을 말할 확률은 0.75이다. 그리고 경험적으로 거짓말 탐지기의 작동이 맞을 확률은 0.85이고 틀릴 확률은 0.15이다. (사후확률 추정) 거짓말 탐지지가 거짓말이라고 응답했을 때, 실제로 이 학생이 진실을 말했을 확률은 얼마인가 ?

거짓말(0.25)		진실(0.75)	
0.85	0.15	0.15	0.85

빨간색은 거짓말 탐지기가 거짓으로 판단할 확률

1. 거짓말, 진실 확률 (사전확률)

A_i, i = A 거짓말 1, 진실2 $P(A_1)=0.25$, $P(A_2)=0.75$

2. 거짓말 탐지기가 거짓일 확률 $P(B)$

$P(B) = \sum_{i=1}^{2} P(A_i)\cdot P(B|A_i)$, 전확률 공식

$0.25\times 0.85 + 0.75\times 0.15 = 0.325$

$\therefore 0.33$

3. 거짓말 중(B), 진실을 말했을 확률 $P(A_2|B)$

$$P(A_2|B) = \frac{P(B|A_2)\cdot P(A_2)}{P(B)} = \frac{0.75\times 0.15}{0.33} = \frac{0.1125}{0.33} = 0.34$$

예3) 한 대학교 재학생의 35%는 여학생이다. (사전정보 확률) 이 학교의 남학생의 70%, 여학생의 20%가 흡연을 한다고 조사되었다. (사후정보 추정) 이 학교의 학생 한 명을 임으로 선발하였을 때, 선출된 학생이 흡연자라면, 이 학생이 여학생일 확률은 얼마인가?

남학생(0.65)		여학생(0.35)	
0.7	0.3	0.2	0.8

빨간색은 흡연자일 확률

1. 학생을 임의로 선택할 확률 (사전확률)

A_i, i = 남학생 1, 여학생 2 $\quad P(A_1) = 0.65,\ P(A_2) = 0.35$

2. 학생 중 흡연자일 확률 $P(B)$

$P(B) = \sum_{i=1}^{2} P(A_i)\cdot P(B|A_i)$, 전확률공식

$0.65 \times 0.7 + 0.35 \times 0.2 = 0.525$

$\therefore\ 0.53$

3. 흡연하는 학생 중 (B), 그 학생이 여학생일 확률 $(A_2|B)$

$$P(A_2|B) = \frac{P(B|A_2)\cdot P(A_2)}{P(B)} = \frac{0.2 \times 0.35}{0.53} = \frac{0.07}{0.53} = 0.132$$

[4] 어느 동아리의 회원 수가 20명이라 하자. 이 중 8명은 안경을 쓰고 있다. 회원들 중 4명을 임의로 추출한다고 했을 때, 4명 중 안경을 쓴 학생의 수를 확률변수 X라 하자.

1) 확률변수 X의 확률질량함수를 정의하고, 확률분포를 구하여라.

2) 확률변수 X의 평균과 분산을 각각 구하여라.

$P(X=x) = {}_8C_x \times {}_{12}C_{4-x} \,/\, {}_{20}C_4$, $x = 0,1,2,3,4$

x	0	1	2	3	4	계
P(X=x)	$\frac{495}{4845}$	$\frac{1760}{4845}$	$\frac{1848}{4845}$	$\frac{672}{4845}$	$\frac{70}{4845}$	1

$\mu = E(X) = \sum_{x=0}^{4} x \cdot P(x)$ ~ 평균을 구하여도 되며, 초기하분포의 평균 =$np = 4 \times \frac{8}{20} = 1.6$

$\sigma^2 = V(X) = E(X^2) - E(X)^2$ ~ 분산을 구하여도 되며, 초기하분포의 분산=$np(1-p)\frac{N-n}{N-1}$

$$= 4 \times \frac{8}{20} \times \left(1 - \frac{8}{20}\right) \times \frac{16}{19} = 0.808$$

=> 확률분포표를 이용해 구한 경우와 근사적인 위 방법으로 구한 경우 수치에 다소 차이가 날 수 있다.

[5] 원통 안에 1에서 10까지 숫자가 쓰여있는 공이 있고 3개를 추첨하여 모두 맞으면 1등, 2개를 맞추면 2등, 1개를 맞추면 3등이 정해진다. 1과 10사이에서 임의의 숫자 3개를 고른 후 추첨에서 3개의 공의 임의로 뽑을 때 2개 이상 맞출 확률을 구하여라.

1) 추첨방식이 복원방법일 경우 확률을 구하라.

2) 추첨방식이 비복원방법일 경우 확률을 구하라.

3) 1등은 10만원, 2등은 5만원, 3등은 1만원의 상금이 주어진다면, 기대수익은 얼마이겠는가?

답안)

복원추출 확률분포 적용 : $X \sim B(n,p) = B(3, \dfrac{3}{10})$, $x = 0,1,2,3$

비복원추출 확률분포 적용 : $X \sim \ HG(N,n,D,x) = HG(10,3,3,x)$, $x = 0,1,2,3$

$B(3, \dfrac{3}{10}) \ \sim \ P(X = x) = {}_3C_x (\dfrac{3}{10})^x (1 - \dfrac{3}{10})^{3-x}$

$HG(10,3,3,x) \sim \ P(X = x) = {}_3C_x \cdot {}_7C_{3-x} / {}_{10}C_3$

이항분포 : $P(X \geq 2) = 0.216$ <= $= {}_3C_2 (\frac{3}{10})^2 (1 - \frac{3}{10})^1 + {}_3C_3 (\frac{3}{10})^3 (1 - \frac{3}{10})^0$

초기하분포 : $P(X \geq 2) = {}_3C_2 \cdot {}_7C_1 / {}_{10}C_3 + {}_3C_3 \cdot {}_7C_0 / {}_{10}C_3 = \dfrac{21}{120} + \dfrac{1}{120} = \dfrac{22}{120} = 0.183$

초기하분포 ~ 1등 확률 : P(3)=1/120, 2등 확률 : P(2)=21/120, 3등 확률 : P(1)=63/120

이항분포 ~ 1등 확률 : P(3) = 0.027, 2등 확률 : P(2) = 0.189, 3등 확률 : P(1) = 0.432

독립사건 : $E(X) = 100000 \times 0.027 + 50000 \times 0.189 + 10000 \times 0.432 = 16470$ 원

종속사건 : $E(X) = 100000 \times 0.0083 + 50000 \times 0.175 + 10000 \times 0.525 = 14830$ 원