

과제 #1

학번: 2014110374

이름 : 정상만

1. 통계학교를 다니고 있는 학생 14명을 대상으로 다음과 같은 자료를 얻었다. 다음 물음에 답하시오.

학생	성별	가족수	출신지역	지지정당	중간성적	기말성적
A	여자	4	전라도	A	70	65
B	여자	5	서울	B	75	80
C	남자	3	경기도	C	80	75
D	여자	3	충청도	A	95	90
E	남자	4	경상도	A	82	72
F	남자	4	전라도	B	86	77
G	여자	5	충청도	C	77	68
H	남자	5	경상도	B	83	83
I	남자	6	충청도	B	92	88
J	여자	5	서울	C	91	79
K	남자	3	경기도	C	87	85
L	여자	4	경기도	A	69	77
M	여자	3	전라도	A	88	90
N	남자	3	서울	C	73	65

(1) 성별에 대한 도수분포표를 만드시오.(도수, 상대도수 포함). 그리고 결과에 대해 설명하시오.

성별	도수	상대도수
남자	7	7/14
여자	7	7/14
합계	14	1

< 설 명 > 남녀 각각의 도수가 같고, 상대도수가 같다. 이는 총원에서 남녀 구성원이 비율이 완벽히 대칭을 이루고 있음을 보여준다. 또한 상대도수의 합이 1이므로 이 도수분포표는 잘 제시되었다.

(2) 지지정당과 출신지역에 대한 분할표를 만들고 결과에 대해 설명하시오.

출신지역	지지정당			합계
	A	B	C	
서울		1	2	3
경기도	1		2	3
충청도	1	1	1	3
전라도	2	1		3
경상도	1	1		2
합계	5	4	5	14

과제 #1

< 설 명 > 서울과 경기도지역 출신 학생들은 C 정당을 지지하는 비율이 높았고, 충청도 출신 학생들은 지지하는 정당의 비율이 고르게 퍼져있다. 전라도 출신 학생들은 A 정당을 지지하는 비율이 높았고, 경상도는 A 정당과 B 정당에 고르게 퍼져있다. 그러나 C 정당은 두 지역 출신 모두 선택하지 않았다. 분할표에 따르면, 전체적으로 B 정당의 지지율이 나머지 정당보다 약하다는 것을 알 수 있고, A 정당과 C 정당은 동일한 지지율을 갖는다. 각 정당별 총 합계와 지역별 합계 모두 14이므로 분할표는 잘 제시되었다.

(3) 중간성적에 대한 도수분포표를 만드시오. (계급의 수 6, 시작점 65.5)

계급	도수	상대도수
65.5 ~ 70.5	1	1/14
70.5 ~ 75.5	3	3/14
75.5 ~ 80.5	2	2/14
80.5 ~ 85.5	2	2/14
85.5 ~ 90.5	3	3/14
90.5 ~ 95.5	3	3/14
합계	14	1

(4) 다음 물음에 답하시오.

① 중간고사의 줄기-잎 그림을 그리시오.

줄기	잎
6	9
7	0 3 5 7
8	0 2 3 6 7 8
9	1 2 5

과제 #1

② 중간고사와 기말고사 평균과 분산을 구하시오.

중간고사	평균	중간평균 $\overline{x_m} = \frac{69+70+73+\cdots+95}{14} = 82$ 점
	분산	중간분산 $(s_m)^2 = \frac{1}{14-1} \left[\sum_{i=1}^{14} x_i^2 - \frac{1}{14} \left(\sum_{i=1}^{14} x_i \right)^2 \right] = 69.2$ (x_i : 점수의 오름차순에 대응)
기말고사	평균	기말평균 $\overline{x_f} = \frac{65+65+68+\cdots+90}{14} = 78.1$ 점
	분산	기말분산 $(s_f)^2 = \frac{1}{14-1} \left[\sum_{i=1}^{14} y_i^2 - \frac{1}{14} \left(\sum_{i=1}^{14} y_i \right)^2 \right] = 73.2$ (y_i : 점수의 오름차순에 대응)

③ 중간고사와 기말고사의 변동계수를 구하시오.

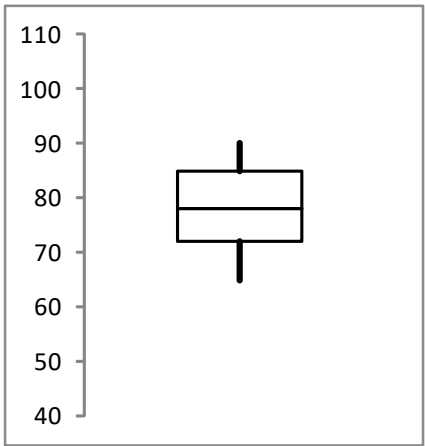
중간고사 변동계수	중간 변동계수 $CV_m = \frac{s_m}{x_m} = \frac{\sqrt{69.2}}{82} \times 100\% = 10.14\%$
기말고사 변동계수	기말 변동계수 $CV_f = \frac{s_f}{x_f} = \frac{\sqrt{73.2}}{78.1} \times 100\% = 10.95\%$

④ 기말고사 사분위수를 구하시오.

제1사분위수: Q_1	$14 \times 0.25 = 3.5 \Rightarrow 3+1$ 번째 값이 제 1사분위수 : 72
제2사분위수: Q_2	$14 \times 0.50 = 7 \Rightarrow 7$ 번째 값과 $7+1$ 번째 값의 평균이 제 2사분위수 : $\frac{77+79}{2} = 78$
제3사분위수: Q_3	$14 \times 0.75 = 10.5 \Rightarrow 10+1$ 번째 값이 제 3사분위수 : 85

과제 #1

⑤ 기말고사의 상자그림을 그리시오.



2. 다음은 어느 도시의 임금 근로자들의 부양 가족수에 대한 확률분포이다. 이 확률분포의 기대값과 표준편차를 구하시오.

부양가족수	0	1	2	3	4	5
확률	0.1	0.2	0.4	0.2	0.05	0.05

$X = \{0, 1, 2, 3, 4, 5\}$ 일 때

기댓값은 $E(X) = 0 \times 0.1 + 1 \times 0.2 + 2 \times 0.4 + 3 \times 0.2 + 4 \times 0.05 + 5 \times 0.05 = 2.05$

$(E(X))^2 = 4.2025$, $E(X^2) = 5.65$ 이므로

분산 $Var(X) = E(X^2) - (E(X))^2 = 5.65 - 4.2025 = 1.4475$ 이고,

따라서 표준편차는 $sd(X) = \sqrt{Var(X)} = \sqrt{1.4475} = 1.2031$

3. 어느 회사에서 사원을 8명 모집하기 위해 신문에 광고를 내었더니 15명의 남자와 8명의 여자가 지원하였다.

(1) 임의로 8명을 선택한다고 한다면 8명 중 여자가 3명일 확률을 구하라.

X : 8명 중 여자 신입사원의 수라 하면 $N = 23, D = 8, n = 8$ 인 초기하분포를 따르므로

$$P(X = 3) = \frac{\binom{8}{3} \binom{23-8}{8-3}}{\binom{23}{8}} = 0.343$$

또는

A : 8명 중 3명이 여자일 사건,

과제 #1

$$P(A) = \frac{\binom{15}{5}\binom{8}{3}}{\binom{23}{8}} = \frac{\frac{15!}{5!(15-5)!} \left(\frac{8!}{3!(8-3)!} \right)}{\frac{23!}{8!(23-8)!}} = \frac{3003(56)}{490314} = \frac{168168}{490314} = 0.343$$

(2) 예상되는 여자 신입사원은 몇 명이 되겠는가?

초기하분포의 기댓값을 구하자.

성공 확률 $p = \frac{D}{N} = \frac{8}{23}$ 이므로 기댓값 $E(X) = np = 8 \times \frac{8}{23} = 2.782$ 이다.

∴ 예상되는 여자 신입사원은 2.782명, 즉 적어도 2명 이상 혹은 3명이 된다.

또는

여자 신입사원이 없을 경우부터 8명 모두 여자 신입사원일 경우를 고려한다. 즉 다음과 같은 확률 모형을 구한 다음 기댓값을 구한다.

$$\begin{aligned} f(0) &= \frac{\binom{15}{8}}{\binom{23}{8}} = 0.013, \quad f(1) = \frac{\binom{15}{7}\binom{8}{1}}{\binom{23}{8}} = 0.105, \quad f(2) = \frac{\binom{15}{6}\binom{8}{2}}{\binom{23}{8}} = 0.286, \\ f(3) &= \frac{\binom{15}{5}\binom{8}{3}}{\binom{23}{8}} = 0.343, \quad f(4) = \frac{\binom{15}{4}\binom{8}{4}}{\binom{23}{8}} = 0.195, \quad f(5) = \frac{\binom{15}{3}\binom{8}{5}}{\binom{23}{8}} = 0.052 \\ f(6) &= \frac{\binom{15}{2}\binom{8}{6}}{\binom{23}{8}} = 0.006, \quad f(7) = \frac{\binom{15}{1}\binom{8}{7}}{\binom{23}{8}} = 0, \quad f(8) = \frac{\binom{8}{8}}{\binom{23}{8}} = 0 \end{aligned}$$

확률함수의 합계는 $f(0) + f(1) + \dots + f(8) = 1$. 따라서 확률모형은

$X = x_i$	0	1	2	3	4	5	6	7	8
$f(x_i)$	0.013	0.105	0.286	0.343	0.195	0.052	0.006	0	0

기댓값 $E(X) = 0.105 + 2 \times 0.286 + 3 \times 0.343 + 4 \times 0.195 + 5 \times 0.052 + 6 \times 0.006 = 2.782$

∴ 예상되는 여자 신입사원은 2.782명, 즉 적어도 2명 이상 혹은 3명이 된다.

과제 #1

4. 대도시의 시민 중 30%만이 대중교통에 만족한다고 한다. 만약 20명의 시민을 임의로 선택하였을 때, 5명 이하의 시민이 대중교통에 만족할 확률과 6명의 시민이 만족할 확률을 구하시오.

대도시의 시민 중 30%만이 대중교통에 만족하므로 각 시민들이 대중교통에 만족할 확률이 0.3이라 하고 X : 20명 중 대중교통에 만족한 시민 수 라고 할 때 이것은 $n = 20$, $p = 0.3$ 인 이항분포를 따르게 된다. 이항분포의 확률 질량함수는 n 이 커짐에 따라 계산이 번거로우므로, 이항분포표를 이용하여 확률을 구하고자 한다.

(i) $P(X \leq 5)$

$n = 20, x \leq 5, p = 0.3$ 이므로 이항분포표에 따르면

20명 중 5명 이하의 시민이 대중교통에 만족할 확률은 $P(X \leq 5) = 0.416$ 이다.

(ii) $P(X = 6)$

위와 동일한 상황에서 $x = 6$ 인 경우만 고려하면 된다.

따라서 $P(X = 6) = P(X \leq 6) - P(X \leq 5) = 0.608 - 0.416 = 0.192$ 이다.

이를 이항분포의 확률질량함수를 이용하여 구하여도 동일한 결과를 얻을 수 있다.

$$P(X = 6) = \binom{20}{6} (0.3)^6 (1 - 0.3)^{20-6} = 38760 (0.000729) (0.006782) = 0.192$$

5. 어느 출판사로부터 출간된 책들에서는 1쪽당 1개꼴로 오타가 생긴다고 알려져 있다. 이 출판사에서 출간한 통계학 책이 300쪽으로 구성되어 있을 때 어느 한쪽에 5개 이상의 오타가 나올 확률은?

X : 300쪽에서 발생하는 오타의 수라고 하면 이는 포아송분포의 가정을 만족하므로

$P(X \geq 5)$ 을 구하면 된다. 평균 1쪽당 1개꼴로 오타가 발생하므로 $m = \frac{300}{300} = 1$ 이된다.

따라서 구하는 확률은 포아송분포표를 이용하면 다음과 같다.

$$P(X \geq 5) = 1 - P(X \leq 4) = 1 - 0.999 = 0.001$$

\therefore 300쪽으로 구성된 책에서 어느 한쪽에 5개 이상의 오타가 나올 확률은 0.001이다.

6. 확률변수 X 와 Y 에 대한 수치들이 다음과 같이 주어져 있다.

$$\mu_X = 2, \sigma_X = 3, E(XY) = 15, \mu_Y = 4, \sigma_Y = 3$$

두 확률변수 X 와 Y 의 공분산과 상관계수를 구하시오.

$$\text{공분산 } Cov(X, Y) = E(XY) - \mu_X \mu_Y = 15 - 2 \times 4 = 7$$

$$\text{상관계수 } Corr(X, Y) = \frac{Cov(X, Y)}{\sigma_X \sigma_Y} = \frac{7}{3 \times 3} = \frac{7}{9} = 0.7778$$

과제 #1

7. 두 확률변수 X 와 Y 의 결합확률분포가 다음과 같이 주어져 있다. 다음 물음에 답하시오.

$x \backslash y$	0	1	2
0	0.05	0.25	0.05
1	0.25	0.25	0.15

$$(1) P[X = Y] = f(0,0) + f(1,1) = 0.05 + 0.25 = 0.3$$

$$P[X > Y] = f(1,0) = 0.25$$

(2) 확률변수 X 와 Y 의 공분산과 상관계수를 구하시오.

X, Y 의 주변확률분포를 우선 구하자.

$$f_X(0) = f(0,0) + f(0,1) + f(0,2) = 0.35, \quad f_X(1) = f(1,0) + f(1,1) + f(1,2) = 0.65$$

$$f_Y(0) = f(0,0) + f(1,0) = 0.3, \quad f_Y(1) = f(0,1) + f(1,1) = 0.5, \quad f_Y(2) = f(0,2) + f(1,2) = 0.2$$

$$\text{그러면 } \mu_X, \mu_Y \text{ 는 } \mu_X = 0.65, \mu_Y = 0.9 \text{ 이고, } E(XY) = f(1,1) + 2 \times f(1,2) = 0.55$$

$$\text{따라서 공분산 } Cov(X, Y) = 0.55 - (0.65)(0.9) = -0.035$$

X, Y 의 표준편차를 구하면

$$\sigma_X = \sqrt{Var(X)} = \sqrt{1 \times 0.65 - (0.65)^2} = \sqrt{0.2275} = 0.477$$

$$\sigma_Y = \sqrt{Var(Y)} = \sqrt{(1 \times 0.5 + 4 \times 0.2) - (0.9)^2} = \sqrt{0.49} = 0.7$$

$$\text{따라서 상관계수 } Corr(X, Y) = \frac{Cov(X, Y)}{\sigma_X \sigma_Y} = \frac{-0.035}{0.477 \times 0.7} = -0.1048$$

(3) 확률변수 X 와 Y 는 서로 독립인가? 이유를 간단히 쓰시오.

공분산과 상관계수 모두 0이 아니므로 독립이 아니다

8. 의사들이 참고로 사용하는 어린이 성장표에 따르면 생후 2년 된 남아의 키는 평균이 88cm이고 표준편차가 3.3cm인 정규분포를 따른다고 한다. 만약 임의로 생후 2년 된 한 남아를 선택하였을 때, 키가 85.3cm에서 93.2cm사이일 확률은 얼마인가?

$N(88, (3.3)^2)$ 일 때

$$P(85.3cm < x < 93.2cm) = P\left(\frac{85.3 - 88}{3.3} < Z < \frac{93.2 - 88}{3.3}\right) = P(-0.82 < Z < 1.58)$$

$$P(-0.82 < Z < 1.58) = 0.9429 - 0.2061 = 0.7368$$

과제 #1

9. 구급차가 병원에서 시내에 있는 스포츠센터까지 가는데 걸리는 시간은 평균이 17분이고 표준편차가 3분인 정규분포를 따른다고 한다.

(1) 소요되는 시간에 대해 다음의 확률을 구하라.

(i) 22분 이상이 소요될 확률

$$N(17, 3^2) \Rightarrow Z = \frac{x-17}{3} \sim N(0,1).$$

$$P(x \geq 22) = P(Z \geq \frac{22-17}{3}) = P(Z \geq 1.67) = P(Z \leq -1.67) = 0.0475$$

(ii) 13분에서 21분 사이의 시간이 소요될 확률

$$P(13 < x < 21) = P(\frac{13-17}{3} < Z < \frac{21-17}{3}) = P(-1.33 < Z < 1.33) = 0.8164$$

(iii) 15.5분에서 18.5분 사이의 시간이 소요될 확률

$$P(15.5 < x < 18.5) = P(\frac{15.5-17}{3} < Z < \frac{18.5-17}{3}) = P(-0.5 < Z < 0.5) = 0.6915 - 0.3085 = 0.3830$$

(2) 도착하는 확률이 가장 높은 1분의 구간은 어디인가?

10. 어느 회사에 입사를 희망한 지원자의 영어점수는 평균이 700이고 표준편차가 100인 정규분포를 따른다고 한다. 다음 물음에 답하여라.

(1) 한 지원자의 점수가 850점 이상의 점수를 받을 확률을 구하여라.

$$P(X \geq 850) = P(Z \geq \frac{850-700}{100}) = P(Z \geq 1.5) = P(Z \leq -1.5) = 0.0668$$

(2) 50%의 지원자를 합격시키기 위해서는 커트라인은 몇 점이겠는가?

커트라인이 x 점이고 x 점 이상인 사람들만 뽑으며, 뽑은 수가 50%이면

$P(X \geq x) = 0.5$ 을 만족시켜야 한다. 우선 $P(Z \geq z) = 0.5$ 을 만족하는 z 를 구하면

$$\text{표준정규분포표로부터 } P(Z \geq 0) = 0.5 \Rightarrow P(\frac{X-700}{100} \geq 0) = P(X \geq 700) = 0.5.$$

$\therefore x = 700$ 점이 커트라인이다.