김은영

중간고사 대체 과제 풀이

1. **[2점]** 표본공간이 ω_1, ω_2, ω_3, ω_4, ω_5로 이루어져 있을 때, ω_1, ω_4, ω_5는 같은 확률을 가지고, ω_2는 ω_1의 두 배의 확률을 가지며, ω_3는 ω_1의 네 배의 확률을 갖는다고 한다.

(1) 각각의 $P(w_i)$를 구하라.

【풀이】

$P(w_1) = P(\omega_4) = P(\omega_5) = p$, $P(w_2) = 2P(\omega_1) = 2p$, $P(w_3) = 4P(\omega_1) = 4p$

$\rightarrow \sum_{i=1}^{5} P(\omega_i) = 9p = 1$, $p = \frac{1}{9}$

$\Rightarrow P(w_1) = P(\omega_4) = P(\omega_5) = p = \frac{1}{9}$, $P(w_2) = 2p = \frac{2}{9}$, $P(w_3) = 4p = \frac{4}{9}$

(2) $A = \{\omega_1, \omega_4\}$일 때, $P(A)$를 구하라.

【풀이】 $P(A) = P(\omega_1) + P(\omega_4) = \frac{1}{9} + \frac{1}{9} = \frac{2}{9}$

6. **[2점]** 중심극한정리(Central Limit Theorem)에 대해 설명하고 이 정리를 이용하면 어떤 이점이 있는지 자신의 의견을 기술하시오.

【풀이】

■ 중심극한정리에 대한 설명: 모집단의 분포가 무엇이든 관계없이 n이 크면 (통상적으로 n이 30보다 크면) 표본평균 $\overline{X}$의 분포는 근사적으로 평균 μ, 표준편차 $\sigma/\sqrt{n}$인 정규분포를 따른다. 따라서 $Z = \frac{\overline{X} - \mu}{\sigma/\sqrt{n}}$은 근사적으로 표준정규분포 $N(0, 1^2)$을 따른다.

■ 이 정리를 사용할 경우의 이점

- 모집단의 분포를 알지 못하는 경우에도 표본수가 충분히 크다면 중심극한정리에 의해 표본의 분포는 정규분포를 따른다고 가정할 수 있으며, 이러한 정규성 가정을 통해 다양한 확률분포를 통계검정에 이용할 수 있다.
- (예) 중심극한정리에 의해서 두 집단의 측정치가 정규분포를 띤다고 가정할 수 있다면 그 평균을 비교함으로써 두 집단의 차이를 쉽게 비교할 수 있다.

2. [2점] 우리나라 사람의 남자 4.3%, 여자 3.6%가 왼손잡이라고 한다. 임의로 남자 500명과 여자 400명을 선정했을 때, 왼손잡이 남자와 여자의 비율의 차이가 0.5% 이하일 확률을 구하라.

【풀이】

남자의 수: n, 여자의 수: m

남자의 비율: p_1, 여자의 비율: p_2

$p_1 - p_2 = 0.043 - 0.036 = 0.007$

$$\sqrt{\frac{p_1 q_1}{n} + \frac{p_2 q_2}{m}} = \sqrt{\frac{0.043 \times 0.957}{500} + \frac{0.036 \times 0.964}{400}} = \sqrt{0.00017} = 0.013$$

따라서 두 표본비율 $\hat{p}_1 - \hat{p}_2$은 다음과 같은 근사표준정규분포를 이룬다.

$$\frac{(\hat{p}_1 - \hat{p}_2) - 0.007}{0.013} \approx N(0,1)$$

그러므로 구하고자 하는 근사확률은

$$P(\hat{p}_1 - \hat{p}_2 \le 0.005) = P\left(Z \le \frac{0.005 - 0.007}{0.013}\right) \approx P(Z \le -0.15) = P(Z \ge 0.15)$$
$$= 1 - P(Z < 0.15) = 1 - 0.5596 = 0.4404$$

※ $(\hat{p}_1 - \hat{p}_2)$ 대신 $|\hat{p}_1 - \hat{p}_2|$을 사용하여 풀이한 경우도 정답처리합니다.

7. [2점] $X \sim B(30, 0.2)$일 때, 확률질량함수를 이용한 $P(X=4)$와 연속성을 수정한 근사확률 $P(X=4)$를 구하라.

【풀이】

$X \sim B(30, 0.2)$ → 확률질량함수 $f(x) = \binom{30}{x}(0.2)^x (0.8)^{30-x}, x = 0, 1, 2, \cdots, 30$

(1) 확률질량함수를 이용: $P(X=4) = \binom{30}{4}(0.2)^4 (0.8)^{26} = 0.1325$

(2) 연속성을 수정한 근사확률

$\mu = np = 30 \times 0.2 = 6$, $\sigma^2 = npq = 30 \times 0.2 \times 0.8 = 4.8$ → $X \approx N(6, 4.8)$에 근사

연속성을 수정하면

$$P(X=4) = P(3.5 \le X \le 4.5)$$
$$= P\left(\frac{3.5-6}{\sqrt{4.8}} \le Z \le \frac{4.5-6}{\sqrt{4.8}}\right) = P(-1.14 \le Z \le -0.68)$$
$$= P(0.68 \le Z \le 1.14) = P(Z \le 1.14) - P(Z \le 0.68)$$
$$= 0.8729 - 0.7517 = 0.1212$$

3. **[3점]** 두 줄기-잎 그림의 표현에 있어 줄기가 너무 적으면 한 줄기를 0~1, 2~3, 4~5, 6~7, 8~9처럼 5개로 나눌 수 있다. 이러한 그림을 '다섯 줄기-잎 그림'이라고 한다. 다음은 다섯 줄기-잎 그림이다.

1	89
2	011
2	2233
2	44455
2	667
2	89
3	0

(1) 이 그림으로부터 원자료를 숫자로 표시하시오.【풀이】

18	19	20	21	21
22	22	23	23	24
24	24	25	25	26
26	27	28	29	30

(2) 위 데이터를 대상으로 '상자-수염 그림'을 그리시오. 상자-수염 그림을 구성하는 5개 요약통계량을 계산한 후 상자-수염 그림에 숫자를 함께 표시하시오.

【풀이】

최소값$=18$, 최대값$=30$

$Q_1 = \frac{21+22}{2} = \frac{43}{2} = 21.5$ (5번째 관측값과 6번째 관측값의 평균)

$Q_2 = \frac{24+24}{2} = 24$ (10번째 관측값과 11번째 관측값의 평균) ← 중앙값

$Q_3 = \frac{26+26}{2} = 26$ (15번째 관측값과 16번째 관측값의 평균)

(3) 데이터 분석 측면에서 '줄기-잎 그림'의 장점을 기술하고, 위에서 주어진 데이터는 어떻게 해석할 수 있는지 자신의 의견을 제시하시오.

【풀이】

■ '줄기-잎 그림'의 장점

- 관측값이 순서대로 정열되어 있어 최소값, 중앙값을 포함한 사분위수, 최대값 등을 쉽게 알 수 있다.
- 자료의 분포에 대해 시각적으로 쉽게 알 수 있다.
- 개별 관측값을 분포 속에서 값으로 하나하나 확인할 수 있다.

■ '줄기-잎 그림'을 통한 상기 데이터 구조에 대한 해석

: 전체 데이터는 24~25를 중심으로 좌우대칭으로 분포하고 있다.

4. [2점] 자료집단 [62, 69, 72, 34, 69, 67, 70, 65, 99]에 대한 표본평균과 15% 절사평균을 구하라.

【풀이】

(1) 표본평균: $\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i = \frac{607}{9} \fallingdotseq 67.44$

(2) 15% 절사평균

자료집단 [62, 69, 72, 34, 69, 67, 70, 65, 99]의 자료값들을 순서대로 정렬함

[34, 62, 65, 67, 69, 69, 70, 72, 99]

$0.15 \times 9 = 1.35$ → 양 끝에서 각각 1개씩 자료값을 제외한 자료의 평균을 구함

[62, 65, 67, 69, 69, 70, 72]의 평균을 구함

$\bar{x}_t = \frac{1}{n_t}\sum_{i=1}^{n_t} x_i = \frac{474}{7} \fallingdotseq 67.71$

5. [3점] 이항분포에 대한 다음 물음에 답하라.

(1) 이항분포 $n=3, p=0.6$에서 이항분포표를 작성하라. 【풀이】

x	$f(x)$	$xf(x)$	$x^2 f(x)$
0	$\binom{3}{0}p^0q^3 = 1\times 0.6^0 \times 0.4^3 = 0.064$	0	0
1	$\binom{3}{1}p^1q^2 = 3\times 0.6^1 \times 0.4^2 = 0.288$	0.288	0.288
2	$\binom{3}{2}p^2q^1 = 3\times 0.6^2 \times 0.4^1 = 0.432$	0.864	1.728
3	$\binom{3}{3}p^3q^0 = 1\times 0.6^3 \times 0.4^0 = 0.216$	0.648	1.944
합계	1	1.8	3.96

(2) (1)의 이항분포표로부터 평균과 표준편차를 계산하라.

【풀이】 평균: $\mu = \sum_i x_i f(x_i) = 1.8$

분산: $\sigma^2 = \sum_i x_i^2 f(x_i) - \mu^2 = 3.96 - 1.8^2 = 0.72$

표준편차: $\sigma = \sqrt{\sigma^2} = \sqrt{0.72} = 0.8485$

(3) (2)의 결과를 다음의 공식에서 얻은 결과와 비교해 보라.

평균$=np$, 표준편차$=\sqrt{npq}$

【풀이】
평균 $= np = 3 \times 0.6 = 1.8$, 표준편차 $= \sqrt{npq} = \sqrt{3 \times 0.6 \times 0.4} = 0.85$

8. **[2점]** 다음은 60명을 대상으로 조사한 자료이다. 임의로 한 명을 추출할 때 그가 남성일 사건을 M이라 하고, 추출한 사람이 길거리 금연에 찬성한 사람일 사건을 A라고 할 때, 사건 M과 A가 서로 독립인지 알아보라.

< 남녀별 길거리 금연에 대한 찬반 의견 >

	남자	여자
금연 찬성	24	14
금연 반대	16	6

【풀이】
이 자료에서 남자인 사건(M)과 찬성인 사건(A)가 서로 독립인지 알아보자.

남자이면서 금연을 찬성하는 사람의 비율: $P(M \cap A) = \dfrac{24}{60} = 0.4$

$P(M) = \dfrac{40}{60}$, $P(A) = \dfrac{38}{60}$ $\rightarrow$ $P(M)P(A) = \dfrac{40}{60} \times \dfrac{38}{60} = 0.422$

$\Rightarrow$ $P(M \cap A) \neq P(M)P(A)$

→ 따라서 남자인 사건과 금연 찬성인 사건은 서로 독립이 아니다.

이를 조건부 확률을 이용해 확인해 보자.

남자 중에서 금연 찬성인 사람의 비율 $P(A|M) = \dfrac{24}{40}$

금연에 찬성하는 사람의 비율 $P(A) = \dfrac{38}{60}$

→ $P(A|M) \neq P(A)$ ⇒ 사건 M과 사건 A는 서로 독립이 아니다.

9. **[1점]** μ를 추정하는 데 있어서 90% 오차한계가 2.5가 되기 위하여 요구되는 표본의 크기는 110이라고 알려져 있다. 이 경우 95% 오차한계가 2이 되기 위하여 요구되는 표본의 크기는 얼마인가?

【풀이】

$$n = \left[\frac{z_{\alpha/2}\sigma}{d}\right]^2 \rightarrow \sigma = \frac{\sqrt{n}\,d}{z_{\alpha/2}} = \frac{\sqrt{110} \times 2.5}{1.645} = 15.9393 \quad \text{여기서 } \alpha = 0.1, \frac{\alpha}{2} = 0.05$$

$$n = \left[\frac{z_{\alpha/2}\sigma}{d}\right]^2 = \left[\frac{1.96 \times 15.9393}{2}\right]^2 = 244.0005 \rightarrow n = 245$$

10. **[2점]** 모평균 μ에 대한 95% 신뢰구간의 의미를 설명하시오.

11. **[1점]** 어떤 제품을 검사하는데 검사원이 착오로 양품을 불량품으로, 불량품을 양품으로 잘못 분류할 확률이 각각 0.03, 0.02이라 한다. 실제 불량률이 6%일 때, 검사에 제출한 한 제품이 불량으로 판정될 확률을 구하라.

【풀이】

* A: 어떤 제품이 실제로 불량일 사상
* D: 검사원이 불량으로 판정할 사상

$P(A)=0.06$, $P(A^c)=1-P(A)=1-0.06=0.94$

$P(D|A^c)=0.03$

$P(D^c|A)=0.02$, $P(D|A)=1-P(D^c|A)=0.98$

관심사상 D는 상호배반인 두 사상 $A\cap D$와 $A^c\cap D$를 합성하여

$D=(A\cap D)\cup(A^c\cap D)$

확률의 덧셈법칙 및 곱셈법칙을 이용하면

$P(D)=P(A\cap D)\cup P(A^c\cap D)=P(A)P(D|A)+P(A^c)P(D|A^c)$

$\quad = (0.06\times 0.98)+(0.94\times 0.03)=0.087$

13. **[2점]** $p-$값의 의미를 가설검정 맥락에서 유의수준 α와 연관시켜 설명하시오.

14. **[2점]** 어떤 축구 선수의 페널티킥 성공률은 70%라고 하자. 이 선수가 어느 날 10번 성공할 때까지 페널티킥 연습을 한다고 할 때 12번 이상 시도해야 할 확률을 구하라.

【풀이】

X: 어떤 축구 선수가 페널티킥을 성공하는 사건

p; 이 선수의 페널티킥 성공률. $p=0.7$

$X\sim B(n,p)$

이 선수가 10번 성공할 때까지 페널티킥 연습을 한다고 할 때 12번 이상 시도해야 할 확률

= 11번째 시도까지 페널티킥을 9번 성공할 확률

$$P(X\le 9)=\sum_{x=0}^{9}\binom{11}{9}0.7^9 0.3^2$$

누적이항분포표에서 이 확률을 찾으면 0.887 즉 88.7%

12. **[2점]** 다음은 확률변수 X의 확률함수이다.

$$f(x) = \frac{1}{84}\binom{5}{x}\binom{4}{3-x}, \quad x = 0, 1, 2, 3$$

(1) 확률변수 X가 갖는 값 각각의 확률을 계산하고 나열하라.

【풀이】

x	$f(x)$	$xf(x)$	$x^2f(x)$
0	$\frac{1}{84}\binom{5}{0}\binom{4}{3} = \frac{1}{84} \times 1 \times 4 = \frac{4}{84}$	$0 \times \frac{4}{84} = 0$	$0^2 \times \frac{4}{84} = 0$
1	$\frac{1}{84}\binom{5}{1}\binom{4}{2} = \frac{1}{84} \times 5 \times 6 = \frac{30}{84}$	$1 \times \frac{30}{84} = \frac{30}{84}$	$1^2 \times \frac{30}{84} = \frac{30}{84}$
2	$\frac{1}{84}\binom{5}{2}\binom{4}{1} = \frac{1}{84} \times 10 \times 4 = \frac{40}{84}$	$2 \times \frac{40}{84} = \frac{80}{84}$	$2^2 \times \frac{40}{84} = \frac{160}{84}$
3	$\frac{1}{84}\binom{5}{3}\binom{4}{0} = \frac{1}{84} \times 10 \times 1 = \frac{10}{84}$	$3 \times \frac{10}{84} = \frac{30}{84}$	$3^2 \times \frac{10}{84} = \frac{90}{84}$
합계		$\frac{140}{84} = 1.67$	$\frac{280}{84} = 3.33$

(2) 확률변수 X의 평균과 표준편차를 구하라.

【풀이】

평균: $\mu = \sum_i x_i f(x_i) = 1.67$

분산: $\sigma^2 = \sum_i x_i^2 f(x_i) - \mu^2 = 3.33 - 1.67^2 = 0.5411$

표준편차: $\sigma = \sqrt{\sigma^2} = \sqrt{0.5411} = 0.7356$

15. **[3점]** 확률변수 X와 Y에 대한 다음의 수치들이 주어져 있을 때 그의 공분산과 상관계수를 구하고 이를 해석하라.

$$\mu_X = 3,\ \mu_Y = 4,\ E(X^2) = 16,\ E(Y^2) = 25,\ E(XY) = 10$$

【풀이】

공분산: $cov(X, Y) = E(XY) - \mu_X \mu_Y = 10 - (3 \times 4) = 10 - 12 = -2$

상관계수: $corr(X, Y) = \frac{cov(X, Y)}{\sigma_X \sigma_Y} = \frac{-2}{\sqrt{7}\ \sqrt{9}} = -0.25$

여기서, $\sigma_X^2 = E(X^2) - \mu_X^2 = 16 - 3^2 = 16 - 9 = 7$

$\sigma_Y^2 = E(Y^2) - \mu_Y^2 = 25 - 4^2 = 25 - 16 = 9$

16. **[2점]** 어느 과수원에서 수확한 사과의 무게는 평균이 $341.7g$이고 표준편차가 $31.2g$인 정규분포를 따른다고 한다. 이 과수원에서 수확한 사과의 무게가 $313.4g$보다 작을 확률은 얼마인가?

【풀이】

$X \sim N(341.7, 31.2^2)$, $Z \sim N(0, 1^2)$

$$P(X < 313.4) = P\left(Z < \frac{313.4 - 341.7}{31.2}\right) = P(Z < 0.91) = 0.1814$$

17. **[4점]** 어떤 제품의 무게(단위: g)는 평균이 100이고, 분산이 25라고 한다. 최근 이 제품을 생산하는 공정의 일부를 새로운 설비로 교체하였는데, 새로운 설비도입으로 제품무게의 평균에 변화가 생겼는지를 알아보기 위해 공정으로부터 제품 100개를 뽑아 조사하였더니, $\bar{x} = 100.78$이었다. 분산은 변화하지 않았다고 가정하고, 평균이 변했는지를 유의수준 $\alpha = 0.05$로 검정해 보시오.

※ **가설검정의 모든 단계를 정확하게 기술하시오.**

【풀이】

(1) 가설 설정: $H_0 : \mu = 100$, $H_1 : \mu \neq 100$ (양측검정)

(2) 주어진 정보 요약

: $\bar{x} = 100.78$, $\mu = 100$, $\sigma^2 = 25$, $n = 100$ (대표본), $\alpha = 0.05$

(3) 검정통계량: $Z = \dfrac{\overline{X} - \mu_0}{\sigma / \sqrt{n}} = \dfrac{100.78 - 100}{5 / \sqrt{100}} = 1.56$

(4) 기각역 설정

기각치: $c = \pm z_{\alpha/2} = \pm z_{0.05/2} = \pm z_{0.025} = \pm 1.96$

기각역: $R : |Z| \geq 1.96$

(5) 가설검정의 결론

: 검정통계치 $z = 1.56$이 기각역에 포함되지 않으므로 귀무가설을 기각할 수 없다. 따라서 대립가설이 맞다고 말할 수 없다.

(6) 함의: 평균이 변했다고 말할 수 없다.

18. [3점] 치명적인 자동차 사고의 55%가 음주운전에 의한 것이라는 보고가 있다. 앞으로 5건의 치명적인 자동차 사고가 날 때, 음주운전에 의하여 사고가 발생할 횟수 X에 대하여 다음 확률을 구하라.

(1) 5번 모두 사고가 날 확률

【풀이】

$X \sim B(5, 0.55)$

$P(X=5) = 1 - P(X \leq 4) = 1 - 0.9497 = 0.0503$

(2) 꼭 3번 사고가 날 확률

【풀이】

$P(X=3) = P(X \leq 3) - P(X \leq 2) = 0.7438 - 0.4069 = 0.3369$

(3) 적어도 1번 이상 사고가 날 확률

【풀이】

$P(X \geq 1) = 1 - P(X=0) = 1 - 0.0185 = 0.9815$