

## 통계학 ( 강좌) 중간고사 1 (16:00~18:00)

※ 답안지에 소속, 학번, 이름을 빠짐없이 기록하였는지 확인 후, 다음 물음에 대한 정답을 반드시 풀이 과정과 함께 잘 정리하여 제출하세요. 부정행위 (계산기 부정사용 포함) 적발 시 즉시 퇴실 조치할 것입니다.

※ 답은 소수점 셋째자리까지 쓰시오.

1. (총 15점 : 각 3점, 틀리면 2점 추가 감점) 다음 물음에 대하여 맞으면 O 틀리면 X로 답하시오.

(1) 통계량은 모집단으로부터 얻어진다. ( X )

(2) 확률변수  $X, Y$  에 대하여  $\text{Corr}(2X, -3Y) = (2 \times (-3))\text{Corr}(X, Y)$  이다. ( X )

(3) 왜도가 양수이면 오른쪽으로 긴 꼬리를 가지는 분포이다. ( O )

(4) 공분산의 값이 클수록 더욱 강한 선형관계를 나타내기 때문에, 공분산이 -5인 자료가 공분산이 -3인 자료보다 더 강한 음의 선형관계를 나타낸다. ( X )

(5) 어느 백화점에서 판매하는 시계 100개 중 5개가 불량품이다. 손님이 시계 3개를 사갈 때 불량품을 사갈 확률을 계산하려고 한다. 이 때 불량품의 수를 확률변수  $X$ 라 하면 확률변수  $X$ 가 취할 수 있는 값은 1, 2, 3이다. ( X )

2. (5점) 세 사건  $E, F, G$  에 대하여 다음이 성립함을 보이시오.

$$P(E|F) = P(E|F \cap G)P(G|F) + P(E|F \cap G^c)P(G^c|F)$$

(풀이)

$$\begin{aligned} & P(E|F \cap G)P(G|F) + P(E|F \cap G^c)P(G^c|F) \\ &= \frac{P(E \cap F \cap G)}{P(F \cap G)} \cdot \frac{P(G \cap F)}{P(F)} + \frac{P(E \cap F \cap G^c)}{P(F \cap G^c)} \cdot \frac{P(G^c \cap F)}{P(F)} \\ &= \frac{P(E \cap F \cap G)}{P(F)} + \frac{P(E \cap F \cap G^c)}{P(F)} \\ &= \frac{P((E \cap F) \cap (G \cup G^c))}{P(F)} = \frac{P(E \cap F)}{P(F)} \\ &= P(E|F) \end{aligned}$$

3. (10점) 확률변수  $(X_1, X_2, X_3)$ 에 대하여  $\text{Var}(X_1) = 5$ ,  $\text{Var}(X_2) = 4$ ,  $\text{Var}(X_3) = 7$  이고  $\text{Cov}(X_1, X_2) = 3$ ,  $\text{Cov}(X_1, X_3) = -2$  이며  $X_2$ 와  $X_3$ 가 독립일 때, 다음 물음에 답하시오.

(1) (5점)  $Y_1 = X_1 - 2X_2 + 3X_3$ ,  $Y_2 = -2X_1 + 3X_2 + 4X_3$  의 공분산을 구하시오.

(풀이)

$$\begin{aligned}
Cov(Y_1, Y_2) &= Cov(X_1 - 2X_2 + 3X_3, -2X_1 + 3X_2 + 4X_3) \\
&= -2var(X_1) + 3Cov(X_1, X_2) + 4Cov(X_1, X_3) \\
&\quad + 4Cov(X_2, X_1) - 6var(X_2) - 8Cov(X_2, X_3) \\
&\quad - 6Cov(X_3, X_1) + 9Cov(X_3, X_2) + 12var(X_3) \\
&= -2var(X_1) - 6var(X_2) + 12var(X_3) \\
&\quad + 7Cov(X_1, X_2) - 2Cov(X_1, X_3) + Cov(X_2, X_3) \\
&= 75
\end{aligned}$$

(2) (5점)  $Y_1$ 과  $Y_2$ 의 상관계수를 구하시오.

(풀이)

$$\begin{aligned}
Var(Y_1) &= var(X_1) + 4var(X_2) + 9var(X_3) - 4Cov(X_1, X_2) + 6Cov(X_1, X_3) - 12Cov(X_2, X_3) \\
&= 5 + 16 + 63 - 12 - 12 = 60
\end{aligned}$$

$$\begin{aligned}
Var(Y_2) &= 4var(X_1) + 9var(X_2) + 16var(X_3) - 12Cov(X_1, X_2) - 16Cov(X_1, X_3) + 24Cov(X_2, X_3) \\
&= 164
\end{aligned}$$

$$Corr(Y_1, Y_2) = \frac{75}{\sqrt{60 \cdot 164}} = \frac{75}{99.1967 \dots} = 0.7560 \dots$$

4. (10점) 다음 물음에 답하시오.

(1) (5점)  $Z_1, \dots, Z_5$  가 서로 독립이고 표준정규분포를 따르는 확률변수일 때, 다음 확률을 구하여라.

$$P(Z_1^2 + Z_2^2 \leq 6.37 \times (Z_3^2 + Z_4^2 + Z_5^2))$$

(풀이)

$X_1 = Z_1^2 + Z_2^2 \sim X_{(2)}^2$ ,  $X_2 = Z_3^2 + Z_4^2 + Z_5^2 \sim X_{(3)}^2$ 을 따르므로,  $Y = \frac{X_1/2}{X_2/3} \sim F(2, 3)$  분포를 따른다.

$$\begin{aligned}
&\Rightarrow P(Z_1^2 + Z_2^2 \leq 6.37(Z_3^2 + Z_4^2 + Z_5^2)) = P\left(\frac{X_1}{X_2} \leq 6.37\right) \\
&= P\left(\frac{X_1/2}{X_2/3} \leq 6.37 \times \frac{3}{2}\right) = P\left(\frac{X_1/2}{X_2/3} \leq 9.555\right) = 1 - 0.05 = 0.95
\end{aligned}$$

(2) (5점) 정규분포  $N(0, 2^2)$  를 따르는 것으로 알려진 모집단에서 크기 16인 랜덤포본을 추출하였을 때,  $P(\bar{X} \leq y \cdot S) = 0.95$  가 성립하는  $y$  를 구하여라.

(풀이)

$$T = \frac{\bar{X}}{S/\sqrt{n}} \sim t(n-1) \text{ 분포를 따르므로, } \frac{\bar{X}}{S/\sqrt{16}} = 4 \times \frac{\bar{X}}{S} \sim t(15) \text{이다.}$$

$$\Rightarrow P(\bar{X} \leq y \times S) = P\left(4 \times \frac{\bar{X}}{S} \leq 4y\right) = 0.95$$

$$\Rightarrow P\left(4 \times \frac{\bar{X}}{S} \geq 4y\right) = 0.05$$

$$\Rightarrow 4y = 1.753, \quad y = 0.43825$$

5. (10점) 한 심리학자가 단어암기 방법에 대한 연구를 실시하였다. 한 고등학교 학생들 전체에게 새로운 암기방법으로 훈련을 시킨 후에 30명의 학생을 랜덤하게 골라 100개의 단어를 기억하도록 실험하여 각 학생이 기억한 단어의 수는 다음 표와 같다.

91	64	98	66	83	87	83	86	80	93
83	75	72	79	90	80	90	71	84	68
69	90	84	78	65	80	92	72	76	77

단,  $\sum x_i = 2406$ ,  $\sum x_i^2 = 195,368$  이다

(1) (5점) 줄기-잎 그림을 그려라. 단 잎의 폭을 10으로 하여라.

(풀이)

9	0	0	0	1	2	3	8				
8	0	0	0	3	3	3	4	4	6	7	
7	1	2	2	5	6	7	8	9			
6	4	5	6	8	9						

(2) (5점) 평균, 표준편차, 중앙값, 사분위수범위(IQR)을 구하라.

(풀이)

$$\bar{x} = 80.2$$

$$s = \sqrt{\frac{1}{29} \times (\sum x_i^2 - n \times \bar{x}^2)} \approx 9.110055$$

$$Q_2 = \frac{x_{(15)} + x_{(16)}}{2}, Q_1 = x_{(8)} = 72, Q_3 = x_{(23)} = 87$$

$$IQR = (72, 87)$$

6. (15점) 확률변수  $X, Y$  의 결합확률함수가 다음과 같이 주어졌을 때, 다음 물음에 답하여라.

$$p(x, y) = \begin{cases} \frac{1}{4}, & (x, y) = (-4, 1), (4, -1), (2, 2), (-2, -2) \\ 0, & o.w. \end{cases}$$

(1) (5점)  $Cov(X, Y) = 0$  임을 보이시오.

(풀이)

$$Cov(X, Y) = E[(X - \mu_x)(Y - \mu_y)] = E(XY) - E(X)E(Y) \text{ 이다.}$$

여기서

$$E(XY) = \frac{1}{4} [(-4)(1) + (4)(-1) + (2)(2) + (-2)(-2)] = \frac{1}{4} (-4 - 4 + 4 + 4) = 0$$

$$E(X) = \frac{1}{4} [(-4) + 4 + 2 + (-2)] = 0, \quad E(Y) = \frac{1}{4} [1 + (-1) + 2 + (-2)] = 0 \text{ 이므로}$$

$$Cov(X, Y) = E[(X - \mu_x)(Y - \mu_y)] = E(XY) - E(X)E(Y) = 0 - 0 = 0$$

(2) (5점)  $X, Y$  는 서로 독립인가? 독립이 아닌가? 이유를 들어 설명하시오.

(풀이)

$X, Y$  는 서로 독립이 아니다.

만약  $X, Y$  가 서로 독립이라면 두 확률변수  $X, Y$  의 결합밀도함수  $p(x, y)$  와  $X, Y$  의 주변밀도함수  $p(x), p(y)$  사이에 다음과 같은 관계가 성립하여야 한다.  $p(x, y) = p(x)p(y)$

그러나 문제에서 주어진 확률변수  $X, Y$  의 결합밀도함수  $p(x, y)$  와 각 확률변수  $X, Y$  의 주변밀

도함수  $p(x), p(y)$  사이에는  $p(x, y) = p(x)p(y)$ 의 관계가 성립하지 않는다.

예를 들어,  $X = -4$ 이고  $Y = 1$ 일 때를 살펴보자.

$$P(X = -4) = \frac{1}{4}, \quad P(Y = 1) = \frac{1}{4}$$

$$P(X = -4, Y = 1) = \frac{1}{4}$$

$$P(X = -4)P(Y = 1) = \frac{1}{4} \times \frac{1}{4} = \frac{1}{16} \quad \text{이므로}$$

$$P(X = -4, Y = 1) \neq P(X = -4)P(Y = 1)$$

따라서  $p(x, y) \neq p(x)p(y)$  이므로 두 확률변수  $X, Y$ 는 서로 독립이 아니다.

(3) (5점)  $g(X, Y) = |X - Y| - 1$  일 때,  $E(g(X, Y))$ 를 구하시오.

(풀이)

$(x, y) = (-4, 1), (4, -1), (2, 2), (-2, -2)$ 를  $g(X, Y) = |X - Y| - 1$ 에 대입하여 나타내면 다음의 표와 같다.

$(x, y)$	$(-4, 1)$	$(4, -1)$	$(2, 2)$	$(-2, -2)$
$g(X, Y)$	4	4	-1	-1

위의 표에서 보면 알 수 있듯이,  $g(X, Y) = |X - Y| - 1$ 은 -1 또는 4, 두 가지 경우를 가진다. 이를 확률분포표로 나타내면 아래와 같다.

$g(X, Y)$	-1	4
$P(g(X, Y))$	$\frac{1}{2}$	$\frac{1}{2}$

따라서

$$E(g(X, Y)) = (-1)\left(\frac{1}{2}\right) + (4)\left(\frac{1}{2}\right) = \left(-\frac{1}{2}\right) + 2 = \frac{3}{2}$$

7. (20점) 다음은 어느 다섯 가족이 소유하고 있는 애완동물의 숫자를 나타낸 자료이다

가족	A	B	C	D	E
애완동물 수	2	6	4	3	1

(1) (5점) 이 다섯 가족을 모집단이라고 할 때, 가족이 소유한 애완동물 수의 모평균과 모분산을 구하여라.

(풀이)

$$\textcircled{1} \text{ 가족이 소유한 애완동물 수의 모평균 : } \mu = \frac{1}{5} \sum_x x = \frac{1}{5}(2 + 6 + 4 + 3 + 1) = \frac{16}{5} = 3.2$$

$\textcircled{2}$  가족이 소유한 애완동물 수의 모분산 :

$$\begin{aligned}\sigma^2 &= \sum_x (x - \mu)^2 p(x) \\ &= \frac{1}{5} \sum_x x^2 - \mu^2 = \left[ \frac{1}{5} (2^2 + 6^2 + 4^2 + 3^2 + 1^2) \right] - (3.2)^2 = \frac{74}{25} = 2.96\end{aligned}$$

(2) (5점) 위의 유한 모집단에서 크기가 2인 표본을 비복원 단순임의추출로 뽑으려고 한다.  
이 때 표본의 확률분포표를 작성하시오.

(풀이)

위의 유한 모집단에서 크기가 2인 표본을 비복원 단순임의추출로 뽑을 때, 가능한 경우의 수는  ${}_5C_2 = 10$ 가지이다. 이 때 표본의 확률분포표는 다음과 같다.

$(x, y)$	A, B (2, 6)	A, C (2, 4)	A, D (2, 3)	A, E (2, 1)	B, C (6, 4)	B, D (6, 3)	B, E (6, 1)	C, D (4, 3)	C, E (4, 1)	D, E (3, 1)
$P(x, y)$	$\frac{1}{10}$	$\frac{1}{10}$	$\frac{1}{10}$	$\frac{1}{10}$	$\frac{1}{10}$	$\frac{1}{10}$	$\frac{1}{10}$	$\frac{1}{10}$	$\frac{1}{10}$	$\frac{1}{10}$

(3) (5점) 이 때, 표본평균의 표본분포표를 작성하시오.

(풀이)

각 표본들의 표본평균은 다음과 같다.

$(x, y)$	A, B (2, 6)	A, C (2, 4)	A, D (2, 3)	A, E (2, 1)	B, C (6, 4)	B, D (6, 3)	B, E (6, 1)	C, D (4, 3)	C, E (4, 1)	D, E (3, 1)
$P(x, y)$	$\frac{1}{10}$	$\frac{1}{10}$	$\frac{1}{10}$	$\frac{1}{10}$	$\frac{1}{10}$	$\frac{1}{10}$	$\frac{1}{10}$	$\frac{1}{10}$	$\frac{1}{10}$	$\frac{1}{10}$
$\bar{X}$	4	3	2.5	1.5	5	4.5	3.5	3.5	2.5	2

이 때, 표본평균의 표본분포표는 다음과 같다.

$\bar{X}$	1.5	2	2.5	3	3.5	4	4.5	5
$P(\bar{X})$	$\frac{1}{10}$	$\frac{1}{10}$	$\frac{2}{10}$	$\frac{1}{10}$	$\frac{2}{10}$	$\frac{1}{10}$	$\frac{1}{10}$	$\frac{1}{10}$

(4) (5점) 표본 평균의 표준오차를 구하시오.

(풀이)

(3)번에서 구한 표본평균의 표본분포표를 이용하면 다음과 같다.

$$\begin{aligned}E(\bar{X}) &= 1.5 \times \frac{1}{10} + 2 \times \frac{1}{10} + 2.5 \times \frac{2}{10} + 3 \times \frac{1}{10} + 3.5 \times \frac{2}{10} + 4 \times \frac{1}{10} + 4.5 \times \frac{1}{10} + 5 \times \frac{1}{10} \\ &= 3.2\end{aligned}$$

$$\begin{aligned}Var(\bar{X}) &= E(\bar{X}^2) - E(\bar{X})^2 \\ &= (1.5^2 \times \frac{1}{10} + 2^2 \times \frac{1}{10} + 2.5^2 \times \frac{2}{10} + 3^2 \times \frac{1}{10} + 3.5^2 \times \frac{2}{10} + 4^2 \times \frac{1}{10} + 4.5^2 \times \frac{1}{10} + 5^2 \times \frac{1}{10}) - 3.2^2 \\ &= 11.35 - 10.24 = 1.11 \\ \therefore SE(\bar{X}) &= \sqrt{Var(\bar{X})} = \sqrt{1.11} = 1.054\end{aligned}$$

또는 유한모집단에서 크기가 2인 표본을 비복원 단순임의추출로 추출하였을 때, 표본평균의 분산

은 유한모집단 수정계수가 곱해진다는 것을 이용하면

$Var(\bar{X}) = \frac{\sigma^2}{n} \left( \frac{N-n}{N-1} \right) = \frac{2.96}{2} \left( \frac{5-2}{5-1} \right) = 1.11$  이므로  $SE(\bar{X}) = \sqrt{1.11} = 1.054$  임을 알 수 있다.

8. (15점) 다음의 확률을 계산하여라.

(1) (5점) 10000명을 대상으로 한 어떤 시험에서 학생들의 성적이 평균  $\mu=18.6$ ,  $\sigma=6.3$  라고 한다. 전체 응시한 학생들 중에서 81명을 랜덤하게 선택할 때, 이들의 평균이 20이상일 확률을 구하여라.

(풀이)

중심극한정리에 의해서 평균 18.6, 분산  $6.3^2$ 인 모집단에서 표본 크기가 81인 임의표본의 표본평균  $\bar{X}$ 는 근사적으로  $N(18.6, \frac{6.3^2}{81})$ 을 따른다. 따라서 표본평균이 20이상일 확률은

$$\begin{aligned} P(\bar{X} \geq 20) &= P\left(\frac{\bar{X}-18.6}{6.3/\sqrt{81}} \geq \frac{20-18.6}{6.3/\sqrt{81}}\right) \\ &= P(Z \geq 2) = 1 - P(Z \leq 2) \\ &= 1 - 0.9772 = 0.0228 \end{aligned}$$

(2) (5점) 서울시는 한강물에 있는 대장균의 분포에 대해 매해 조사하고 있다. 한강물 1ml당 대장균수의 평균과 분산은 각각  $\mu=150$ ,  $\sigma^2=100$ 의 정규분포를 따른다고 하자. 1ml당 대장균수가 155 이상인 지역은 수질 관리대상으로 정하려고 한다. 한강대교 부근 10군데에서 대장균수를 검사했고, 그 평균이 155과 비교되었을 때, 한강대교 부근이 수질 관리대상이 될 확률을 구하여라.

(풀이)

모집단이 정규분포  $N(150, 10^2)$ 을 따를 때 표본 크기가 10인 임의표본의 표본평균  $\bar{X}$ 는  $N(150, \frac{10^2}{10})$ 을 따른다. 따라서 표본평균이 155이상일 확률은

$$\begin{aligned} P(\bar{X} \geq 155) &= P\left(\frac{\bar{X}-150}{10/\sqrt{10}} \geq \frac{155-150}{10/\sqrt{10}}\right) \\ &\approx P(Z \geq 1.58) = 1 - P(Z \leq 1.58) \\ &= 1 - 0.9429 = 0.0571 \end{aligned}$$

(3) (5점) 하루 10,000개의 전구를 생산하는 공장이 있다. 이 공장에서 생산되는 전구의 불량률이 0.003이라 한다. 어느 날 생산한 전구 중 불량품이 20개 이하가 될 확률을 구하여라.

(풀이)

확률변수  $X$ 를  $X$ =생산한 전구 중 불량품의 수로 정의하면  $X$ 는  $B(10000, 0.003)$ 을 따른다. 한편  $X$ 의 평균과 분산은 각각

$$E(X) = np = 10000 \times 0.003 = 30, \quad Var(X) = npq = 10000 \times 0.003 \times 0.997 = 29.91$$

이다. ( $np$ 와  $n(1-p)$ 이 5보다 크므로,)  $X$ 는 근사적으로  $N(30, 29.91)$ 을 따른다. 따라서 불량품이 20개 이하가 될 확률은 다음과 같다.

$$\begin{aligned} P(X \leq 20) &= P\left(\frac{X-30}{\sqrt{29.91}} \leq \frac{20-30}{\sqrt{29.91}}\right) \\ &\approx P(Z \leq -1.83) = P(Z \geq 1.83) = 1 - P(Z \leq 1.83) \\ &= 1 - 0.966 = 0.0336 \end{aligned}$$

F 분포표

$$F_{\alpha} : P(F \geq F_{\alpha}) = \alpha, F \sim F(df_1, df_2)$$

		표준화의 자유도		
표준화의 자유도	$\alpha$	2	3	4
2	.100	9.00	9.16	9.24
	.050	19.00	19.16	19.25
	.025	39.00	39.17	39.25
	.010	99.00	99.16	99.25
3	.100	5.46	5.39	5.34
	.050	9.55	9.28	9.12
	.025	16.04	15.44	15.10
	.010	30.82	29.46	28.71
4	.100	4.32	4.19	4.11
	.050	6.94	6.59	6.39
	.025	10.65	9.98	9.60
	.010	18.00	16.69	15.98

t 분포표

$$t_{\alpha} : P(T \geq t_{\alpha}) = \alpha, T \sim t(df)$$

df \ $\alpha$	0.10	0.05	0.025	0.01
1	3.078	6.31	12.71	31.82
2	1.886	2.920	4.303	6.965
3	1.638	2.353	3.182	4.541
4	1.533	2.132	2.776	3.747
5	1.476	2.015	2.571	3.365
6	1.440	1.943	2.447	3.143
7	1.415	1.895	2.365	2.998
8	1.397	1.860	2.306	2.896
9	1.383	1.833	2.262	2.821
10	1.372	1.812	2.228	2.764
11	1.363	1.796	2.201	2.718
12	1.356	1.782	2.179	2.681
13	1.350	1.771	2.160	2.650
14	1.345	1.761	2.145	2.624
15	1.341	1.753	2.131	2.602
16	1.337	1.746	2.120	2.583
17	1.333	1.740	2.110	2.567
18	1.330	1.734	2.101	2.552
19	1.328	1.729	2.093	2.539
20	1.325	1.725	2.086	2.528

[illegible]