

## 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.

※ 적절한 풀이과정이 없는 경우에는 정답으로 인정하지 않습니다.

※ 소수점 셋째자리까지 쓰세요.

1. (총 10점, 각 2점) 다음 명제에 대하여 참(O) 또는 거짓(X)을 판별하시오.

(1) 확률변수  $X, Y$ 에 대하여  $X \sim N(\mu_1, \sigma^2)$ ,  $Y \sim N(\mu_2, \sigma^2)$  이고 서로 독립이면,

$X+Y \sim N(\mu_1+\mu_2, \sigma^2)$ 이다. ( X )

(2)  $X_1, \dots, X_n \sim i.i.d. N(\mu, 1)$  일 때,  $n \left( \frac{\bar{X} - \mu}{S} \right)^2 \sim F(n-1, 1)$ 이다. ( X )

$$\text{단, } \bar{X} = \frac{1}{n} \sum_{i=1}^n X_i, S^2 = \frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X})^2.$$

(3) 평균  $\mu$ , 분산  $\sigma^2$ 인 정규분포를 따르는 확률 변수를 제공하면 자유도가 1인  $\chi^2$  분포를 따른다. ( X )

(4) 확률변수  $F$ 가 분자, 분모의 자유도가 각각  $k_1, k_2$ 인  $F$  분포를 따를 때,

$$F_\alpha(k_1, k_2) = \frac{1}{F_{1-\alpha}(k_1, k_2)} \quad \text{가 성립한다. ( X )}$$

(5) 자유도가  $k$ 인 카이제곱 분포에서 추출한 랜덤 표본  $X_1, \dots, X_n$ 에 대하여  $n\bar{X}$ 의 분포는 자유도가  $nk$ 인 카이제곱 분포를 따른다. ( O )

2. (총 8점) 초등학교 학생들의 읽기 능력을 파악하기 위하여 3학년 학생들을 대상으로 작성된 읽기능력평가 시험을 시행한 결과 다음과 자료를 얻었다.

97	98	84	88	81	86	87	70	94	77	
66	85	63	68	46	72	59	36	72	80	68

(1) (3점) 줄기-잎 그림을 그리시오.

풀이)

3 | 6

4 | 6

5 | 9

6 | 3688

7 | 0227

8 | 0145678

9 | 478

(2) (3점) 평균, 표준편차, 중앙값을 구하시오.

(풀이) 표본평균:  $\bar{x} = \frac{1}{21}(97+98+\dots+68) = 75.095$

표본 표준편차 :  $s = \sqrt{\frac{1}{n-1}(\sum x_i^2 - n\bar{x}^2)} = \sqrt{\frac{1}{20}(123463 - 21 \times (75.095)^2)} = 15.872$

표본 중앙값 :  $21 \times 0.5 = 10.5$  이므로 중앙값은  $[10.5] + 1 = 11$  번째 값이 된다. (단,  $[n\alpha]$ 는  $n\alpha$ 의 정수부분을 뜻함) 따라서 중앙값은  $\hat{Q}_2 = 77$ .

(3) (2점) 사분위 범위를 구하시오.

$\hat{Q}_1$ :  $[21 \times 0.25] + 1 = 6$ 번째 값이므로  $\hat{Q}_1 = x_{(6)} = 68$

$\hat{Q}_3$ :  $[21 \times 0.75] + 1 = 16$ 번째 값이므로  $\hat{Q}_3 = x_{(16)} = 86$

따라서  $\widehat{IQR} = 86 - 68 = 18$ .

3. (총 6점) 선준이는 이메일 보관함을 “스팸 편지함”, “보통 편지함”, “중요 편지함”, 이렇게 3개의 범주로 구분하여 수신된 이메일을 정리한다. 과거의 경험에 비추어 선준이는 전체 편지 중 70%가 스팸 편지함에 속하고, 20%가 보통 편지함에, 나머지 10%가 중요 편지함으로 분류된다는 것을 확인하였다. 또한, “공짜”라는 단어가 내용에 포함된 이메일들을 체크해 본 결과, 스팸 편지함의 편지들은 90%가 이 단어를 포함하고 있었고, 보통 편지함의 편지들은 1%, 중요 편지함의 편지들 역시 1%만이 이 단어를 포함하고 있었다.

(1) (3점) 선준이가 새롭게 받은 이메일에 “공짜”라는 단어가 포함되어 있을 확률은 얼마인가?

(풀이) 선준이가 새롭게 받은 이메일에 “공짜”라는 단어가 포함되어 있는 사건을  $S$  라고 하고, 이메일이 스팸 편지함, 보통편지함, 중요편지함으로 분류되는 사건을 각각  $A, B, C$  라고 하자.

$$P(S|A)=0.9, P(S|B)=0.01, P(S|C)=0.01$$

$$P(A)=0.7, P(B)=0.2, P(C)=0.1.$$

따라서 구하는 확률은 다음과 같다.

$$P(S) = P(S|A)P(A) + P(S|B)P(B) + P(S|C)P(C) = 0.633$$

(2) (3점) 선준이가 새롭게 받은 이메일에 “공짜”라는 단어가 포함되어 있을 때, 이 편지가 스팸일 확률은 얼마인가?

(풀이)

$$P(A|S) = \frac{P(S|A)P(A)}{P(S)} = \frac{0.9 \times 0.7}{0.633} = 0.995$$

4. (총 9점) 다음 함수들에 대하여 이 함수들을 확률밀도함수로 사용할 수 있는지를 결정하고 그 이유에 대해 설명하시오.

(1) (3점)  $p(x) = \frac{x}{8}, \quad x=3,4,5,6$

(2) (3점)  $p(x) = \frac{x^2}{30}, \quad x=0,1,2,3,4$

(3) (3점)  $p(x) = \frac{5-x^2}{6}, \quad x=0,1,2,3$

(풀이) 확률밀도함수가 되기 위해서는  $p(x) \geq 0, \sum p(x) = 1$ 을 만족해야 한다.

(1)  $p(x) \geq 0$ 의 조건은 만족하지만  $\sum p(x) > 1$ 이므로 확률밀도함수가 될 수 없다.

(2) 두 성질을 모두 만족하므로 확률밀도함수가 된다.

(3)  $\sum p(x) = 1$ 은 만족하지만  $x=3$ 에서  $p(3) < 0$ 이므로 확률밀도함수가 될 수 없다.

5. (총 10점) 다음 물음에 답하여라.

(1) (5점) A와 B 두 과목으로 이루어진 한 시험이 있다. A과목의 성적은 평균이 70점이고, 표준편차가 10점인 정규분포를 따르고, B과목의 성적은 평균이 60점이고 표준편차가 15점인 정규분포를 따르며 두 과목의 성적은 서로 독립이다. 두 과목의 평균이 상위 5%인 학생에게 전액 장학금이 주어진다고 할 때, A과목이 95점인 학생이 전액 장학금을 받기 위해서는 최소 B과목을 몇 점 이상 받아야 하는가?

(풀이)  $A \sim N(70, 10^2), B \sim N(60, 15^2)$

A와 B는 독립이므로  $S = \frac{A+B}{2} \sim N(65, \frac{325}{4})$  이다.

$$P\left[S > \frac{95+b}{2}\right] \leq .05 \Rightarrow P\left[\frac{S-65}{\sqrt{(325/4)}} > \frac{(95+b)/2-65}{\sqrt{(325/4)}}\right] = 0.05$$

$$\Rightarrow \frac{(95+b)/2-65}{9.014} \geq 1.645 \Rightarrow b \geq 64.656$$

따라서 최소 65점 이상을 받아야 한다.

(2) (5점) 어느 회사에 걸려오는 전화 중 30%는 장거리 전화라고 한다. 어느 날 걸려온 100회의 전화 중 장거리 전화가 25회 이하일 확률을 얼마인가?

(풀이)  $X$  : 장거리 전화 횟수  $\sim B(100, 0.3)$

$np > 5, n(1-p) > 5$ 를 만족하므로, 이항분포의 정규근사를 이용하면 :  $X \approx N(30, 21)$

- 연속성 수정 이용하는 경우.

$$P[X \leq 25] \approx P[X \leq 25+0.5] = P\left[\frac{X-30}{\sqrt{21}} \leq \frac{25.5-30}{\sqrt{21}}\right] = P[Z \leq -0.98] = 0.1635$$

- 연속성수정 이용하지 않을 경우.

$$P[X \leq 25] = P\left[\frac{X-30}{\sqrt{21}} \leq \frac{25-30}{\sqrt{21}}\right] \approx P[Z \leq -1.091] \approx 0.1379$$

6. (총 8점) 아래와 같은 분포를 갖는 무한 모집단이 있다. 다음 물음에 답하시오.

$x$	1	2
$p(x)$	0.5	0.5

(1) (4점)  $X_1, X_2, X_3$  가 주어진 모집단에서 추출한 랜덤포본일 때, 새롭게 정의된 확률변수  $Z_1 = \frac{X_1 + X_2 + X_3}{3}$  의 확률분포를 구하고,  $Z_1$ 의 평균과 분산을 계산하여라.

(풀이) 확률분포는 다음과 같다.

$z_1$	1	4/3	5/3	2
$p(z_1)$	1/8	3/8	3/8	1/8

$$E(Z_1) = 1 \times \frac{1}{8} + \left(\frac{4}{3}\right) \times \left(\frac{3}{8}\right) + \left(\frac{5}{3}\right) \times \left(\frac{3}{8}\right) + 2 \times \frac{1}{8} = \frac{3}{2}$$

$$V(Z_1) = E(Z_1^2) - E(Z_1)^2 = \left[1^2 \times \frac{1}{8} + \left(\frac{4}{3}\right)^2 \times \left(\frac{3}{8}\right) + \left(\frac{5}{3}\right)^2 \times \left(\frac{3}{8}\right) + 2^2 \times \frac{1}{8}\right] - \left(\frac{3}{2}\right)^2 = \frac{1}{12}$$

(2) (4점)  $X_1, X_2, \dots, X_{20}$ 이 주어진 모집단에서 추출한 랜덤포본일 때, 표본평균  $\bar{X}$ 의 평균과 분산을 계산하여라.

(풀이) 주어진 모집단의 평균과 분산은 다음과 같다.

$$\mu = 1 \times 0.5 + 2 \times 0.5 = 1.5,$$

$$\sigma^2 = (1^2 \times 0.5 + 2^2 \times 0.5) - 1.5^2 = 0.25$$

따라서 표본평균의 평균과 분산은 각각 다음과 같다.

$$E(\bar{X}) = \mu = 1.5,$$

$$Var(\bar{X}) = \sigma^2/n = 1/80$$

7. (총 15점) 확률변수  $X$ 와  $Y$ 의 결합확률분포가 아래 표와 같다. 다음 물음에 답하시오.

$y$	$x$			$P(Y=y)$
	0	1	2	
-1	㉠	㉡	㉢	1/2
1	㉣	1/2	㉤	1/2
$P(X=x)$	1/6	2/3	1/6	1

(1) (2점) ㉠~㉤를 채워 표를 완성하시오.

(풀이) ㉠ 1/6, ㉡ 0, ㉢ 1/6, ㉣ 1/6, ㉤ 0

$$\begin{aligned} \text{㉡} : P(X=1) &= P(X=1, Y=-1) + P(X=1, Y=1) = P(X=1, Y=-1) + 1/2 = 2/3 \\ \therefore P(X=1, Y=-1) &= 1/6 \end{aligned}$$

$$\text{㉢, ㉤} : P(Y=1) = P(X=0, Y=1) + P(X=1, Y=1) + P(X=2, Y=1) = 1/2 \text{ 이고}$$

$$\text{확률값이 0 이상이어야 하므로 } P(X=0, Y=1) = P(X=2, Y=1) = 0$$

마찬가지로 ㉣, ㉤ 도 계산 가능하다.

(2) (6점)  $E(X)$ ,  $E(Y)$ ,  $E(XY)$ ,  $Var(X)$ ,  $Var(Y)$ ,  $Corr(X, Y)$ 를 구하시오.

(풀이)  $E(X) = 0 \times 1/6 + 1 \times 2/3 + 2 \times 1/6 = 1$ ,  
 $E(Y) = 0$ ,  $E(XY) = 0$

$Var(X) = E(X^2) - E(X)^2 = 1/3$ ,  
 $Var(Y) = 1$

$Corr(X, Y) = \frac{Cov(X, Y)}{\sqrt{Var(X) Var(Y)}} = 0$

(3) (4점) 두 확률변수  $X$ 와  $Y$ 는 서로 독립인가? 그 이유는 무엇인가?

(풀이)  $P(X=0, Y=1) = 0$   
 $P(X=0)P(Y=1) = 1/6 \times 1/2$

모든  $x, y$ 에 대해서  $P(X=x, Y=y) = P(X=x)P(Y=y)$ 가 성립하지 않으므로 독립이 아니다.

(4) (3점)  $Var(X+Y)$ ,  $Var(X-Y)$ 를 구하시오.

(풀이)  $Var(X+Y) = Var(X) + Var(Y) + 2Cov(X, Y) = 4/3$

$Var(X-Y) = Var(X) + Var(Y) - 2Cov(X, Y) = 4/3$

8. (7점)  $X_1, X_2, X_3, X_4, X_5$ 는 정규분포  $N(0, 2^2)$ 을 따르며 서로독립인 확률변수들이다. 새로운 확률변수를 다음과 같이 정의하자.

$$Y = \frac{X_1}{\sqrt{X_2^2 + X_3^2 + X_4^2 + X_5^2}}$$

이 때,  $P(Y > 0.595)$ 를 계산하여라.

( 풀이 )  $\frac{X_1}{2}, \frac{X_2}{2}, \frac{X_3}{2}, \frac{X_4}{2}, \frac{X_5}{2} \stackrel{iid}{\sim} N(0, 1)$ 이므로  $\frac{X_2^2}{4} + \frac{X_3^2}{4} + \frac{X_4^2}{4} + \frac{X_5^2}{4} \sim \chi^2(4)$ 이다.

$$Y = \frac{2 \times \frac{X_1}{2}}{\sqrt{[16 \times (\frac{X_2^2}{4} + \frac{X_3^2}{4} + \frac{X_4^2}{4} + \frac{X_5^2}{4})]/4}} = \frac{2Z}{4\sqrt{V/4}}, \quad Z \sim N(0, 1), \quad V \sim \chi^2(4)$$

$$= \frac{1}{2}T, \quad T \sim t(4)$$

$P(Y > 0.595) = P(T > 1.19) = 0.15$  (  $\because$  자유도가 4인 t분포표 이용 )

9. (7점) 3개의 동전을 동시에 던지는 시행을 20회 반복하여 앞면이 1개, 뒷면이 2개 나오는 횟수를  $X$ 라 할 때, 중심극한정리를 이용하여  $P(X \leq k) < 0.1$ 를 만족하는 가장 큰 양의 정수  $k$ 를 구하시오.

(풀이) 3개의 동전을 동시에 던지는 시행에서 앞면이 1개, 뒷면이 2개가 나타날 확률은  $p = 3/8$ 이다. 따라서 확률변수  $X$ 는  $B(20, 3/8)$ 의 분포를 따르게 되고 확률밀도함수는 다음과 같다.

$$P(X=x) = \binom{20}{x} (3/8)^x (5/8)^{20-x}$$

$np=20 \times (3/8) > 5$ ,  $n(1-p)=20 \times (1-3/8) > 5$ 의 조건을 만족하므로, 중심극한 정리를 이용하면,  $X$ 의 분포가  $N(np, np(1-p))$ 로 근사함을 알 수 있다.

$$\text{즉, } P(X \leq k) = P\left(\frac{X-np}{\sqrt{npq}} \leq \frac{k-np}{\sqrt{npq}}\right) \cong P\left(Z \leq \frac{k-15/2}{\sqrt{75/16}}\right) \text{ where } Z \sim N(0,1) \text{ 를 만족한다.}$$

표준정규분포표를 이용하면,  $z_{0.01} \approx 1.28$  이므로  $\frac{k-15/2}{\sqrt{75/16}} \leq -1.28$ 이어야 함을 알 수 있다.

따라서 식을  $k$ 에 대하여 정리하면,  $k \leq 15/2 - 1.28\sqrt{75/16} = 4.7288$  이 되므로 식을 만족하는 가장 큰 양의 정수  $k$ 는 4이다.

연속성 수정 이용하는 경우에도,  $\frac{k+1/2-15/2}{\sqrt{75/16}} \leq -1.28$ ,  $k \leq 7 - 1.28\sqrt{75/16} = 4.2287$  이므로 이

경우에도 가장 큰 양의 정수  $k$ 는 4이다.

표준 정규 분포표  $P(Z \leq z)$ [illegible]

t 분포표

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

df \ $\alpha$	0.15	0.10	0.05	0.025	0.01
1	1.963	3.078	6.31	12.71	31.82
2	1.386	1.886	2.920	4.303	6.965
3	1.250	1.638	2.353	3.182	4.541
4	1.190	1.533	2.132	2.776	3.747
5	1.156	1.476	2.015	2.571	3.365
6	1.134	1.440	1.943	2.447	3.143
7	1.119	1.415	1.895	2.365	2.998
8	1.108	1.397	1.860	2.306	2.896
9	1.100	1.383	1.833	2.262	2.821
10	1.093	1.372	1.812	2.228	2.764
11	1.088	1.363	1.796	2.201	2.718
12	1.083	1.356	1.782	2.179	2.681
13	1.079	1.350	1.771	2.160	2.650
14	1.076	1.345	1.761	2.145	2.624
15	1.074	1.341	1.753	2.131	2.602
16	1.071	1.337	1.746	2.120	2.583
17	1.069	1.333	1.740	2.110	2.567
18	1.067	1.330	1.734	2.101	2.552
19	1.066	1.328	1.729	2.093	2.539
20	1.064	1.325	1.725	2.086	2.528

$\chi^2$  분포표

$$\chi_{\alpha}^2 : P(\chi^2 \geq \chi_{\alpha}^2) = \alpha, \chi^2 \sim \chi^2(df)$$

df \ $\alpha$	0.975	0.95	0.90	0.10	0.05	0.025
1	0.00	0.00	0.02	2.71	3.84	5.02
2	0.05	0.10	0.21	4.61	5.99	7.82
3	0.22	0.35	0.58	6.25	7.81	9.35
4	0.48	0.71	1.06	7.78	9.49	11.14
5	0.83	1.15	1.61	9.24	11.07	12.83
6	1.24	1.64	2.20	10.64	12.59	14.45
7	1.69	2.17	2.83	12.02	14.07	16.01
8	2.18	2.73	3.49	13.36	15.51	17.53
9	2.70	3.33	4.17	14.68	16.92	19.02
10	3.25	3.94	4.87	15.99	18.31	20.48
11	3.82	4.57	5.58	17.28	19.68	21.92
12	4.40	5.23	6.30	18.55	21.03	23.34
13	5.01	5.89	7.04	19.81	22.36	24.74
14	5.63	6.57	7.79	21.06	23.68	26.12
15	6.26	7.26	8.55	22.31	25.00	27.49
16	6.91	7.96	9.31	23.54	26.30	28.85
17	7.56	8.67	10.09	24.77	27.59	30.19
18	8.23	9.39	10.86	25.99	28.87	31.53
19	8.91	10.12	11.65	27.02	30.14	32.85
20	9.59	10.85	12.44	28.41	31.41	34.17

F 분포표

$$F_{0.05} : P(F \geq F_{0.05}) = 0.05, F \sim F(df_1, df_2)$$

	df_1												
df_2	1	2	3	4	5	6	7	8	9	10	11	12	13
1	161.4	199.50	215.71	224.58	230.16	233.99	236.77	238.88	240.54	241.88	242.98	243.91	244.69
2	18.51	19.000	19.164	19.247	19.296	19.330	19.353	19.371	19.385	19.396	19.405	19.413	19.419
3	10.12	9.552	9.277	9.117	9.013	8.941	8.887	8.845	8.812	8.786	8.763	8.745	8.729
4	7.709	6.944	6.591	6.388	6.256	6.163	6.094	6.041	5.999	5.964	5.936	5.912	5.891
5	6.608	5.786	5.409	5.192	5.050	4.950	4.876	4.818	4.772	4.735	4.704	4.678	4.655
6	5.987	5.143	4.757	4.534	4.387	4.284	4.207	4.147	4.099	4.060	4.027	4.000	3.976
7	5.591	4.737	4.347	4.120	3.972	3.866	3.787	3.726	3.677	3.637	3.603	3.575	3.550
8	5.318	4.459	4.066	3.838	3.687	3.581	3.500	3.438	3.388	3.347	3.313	3.284	3.259
9	5.117	4.256	3.863	3.633	3.482	3.374	3.293	3.230	3.179	3.137	3.102	3.073	3.048
10	4.965	4.103	3.708	3.478	3.326	3.217	3.135	3.072	3.020	2.978	2.943	2.913	2.887
11	4.844	3.982	3.587	3.357	3.204	3.095	3.012	2.948	2.896	2.854	2.818	2.788	2.761
12	4.747	3.885	3.490	3.259	3.106	2.996	2.913	2.849	2.796	2.753	2.717	2.687	2.660
13	4.667	3.806	3.411	3.179	3.025	2.915	2.832	2.767	2.714	2.671	2.635	2.604	2.577
14	4.600	3.739	3.344	3.112	2.958	2.848	2.764	2.699	2.646	2.602	2.565	2.534	2.507
15	4.543	3.682	3.287	3.056	2.901	2.790	2.707	2.641	2.588	2.544	2.507	2.475	2.448
16	4.494	3.634	3.239	3.007	2.852	2.741	2.657	2.591	2.538	2.494	2.456	2.425	2.397
17	4.451	3.592	3.197	2.965	2.810	2.699	2.614	2.548	2.494	2.450	2.413	2.381	2.353
18	4.414	3.555	3.160	2.928	2.773	2.661	2.577	2.510	2.456	2.412	2.374	2.342	2.314
19	4.381	3.522	3.127	2.895	2.740	2.628	2.544	2.477	2.423	2.378	2.340	2.308	2.280