

## 2024년도 1학기 <통계학> 중간고사

-교수 : 정상아-

1. (16점, 각 2점) 다음에 대해 간략하게 설명하시오. (부분점수 없음, 간략하게 핵심적인 개념을 설명할 것)

- 중심극한정리

- 표준오차

- 표준편차

- 제 1종 오류

- 유의수준

- 표본분포

- 편향성

- 상관계수

2. (각 5점, 총 10점) 다음은 50점 만점으로 채점된 어느 시험 점수의 다섯 수치 요약값이다. 아래 물음에 답하시오.

Min=16.5, Q1=32, Median=39, Q3=43.5, Max=48.5

(1) 이 시험의 분포는 [음의 왜도 / 양의 왜도 / 좌우 대칭] 중 어떠한 형태인 것으로 생각되는가? 그 이유는?

(2) 이 시험은 상위 50%만이 통과한다고 할 때, 이 시험의 평균점수를 받은 학생은 시험을 통과할 수 있는가? 그 이유는?

3. (8점) 다음과 같은 분포를 갖는 무한 모집단을 생각해보자.

$$f(x) = \begin{cases} x & \text{for } 0 \leq x \leq 1 \\ 2-x & \text{for } 1 < x \leq 2 \\ 0 & \text{otherwise} \end{cases}$$

(1) (3점) 주어진 모집단에서 선택한 랜덤 표본을  $(X_1, X_2, \dots, X_n)$ 라고 하자.  $X_i$ 의 기댓값과 분산을 구하시오.

(2) (5점) 주어진 모집단에서 크기  $n=40$ 의 랜덤 표본을 선택하여  $Y = \sum_{i=1}^{40} X_i$  로 정의하였다. 이 때,  $P(Y < k) = 0.9$  를 만족하는  $k$ 를 구하시오. 단,  $Y$ 의 분포를 사용하여 답하시오.

4. (총 10점) 빨간색 공 3개와 파란색 공 5개가 들어있는 상자가 있다. 이 상자에서 2개의 공을 랜덤하게 선택하였을 때, 빨간공의 개수를 확률변수  $X$  로 정의하자.

(1) (2점) 비복원 추출로 2개의 공을 랜덤하게 선택하였을 때, 2개 중에서 빨간색 공이 나오지 않을 확률은 무엇인가? 이 때 사용한  $X$  의 분포는 무엇인가?

(2) (2점) 복원추출로 2개의 공을 랜덤하게 선택하였을 때, 2개 중에서 빨간색 공이 나오지 않을 확률은 무엇인가? 이 때 사용한  $X$  의 분포는 무엇인가?

(3) (2점) (1)과 (2)에서  $X$ 의 분포는 서로 동일한가? 그렇다면 혹은 그렇지 않다면 그 이유는 무엇인지 간략하게 쓰시오.

(4) (4점) 공을 더 추가하여 빨간색 공 30개와 파란색 공 50개가 든 상자에서 비복원 추출로 2개의 공을 랜덤하게 선택하였을 때, 2개 중에서 빨간색 공이 나오지 않을 확률은 무엇인가?

5. (총 10점) 어느 기관의 조사에 따르면, 1년 동안 발신되는 이메일의 25%가 스팸 메일이라고 한다. 이 기관에서 발신하는 이메일 중 일부를 랜덤하게 선택했을 때, 이 중 조사된 스팸 메일의 수를  $X$  라고 하자. 단, 발신된 이메일은 서로 독립적이라고 한다.

(1) (5점) 60통의 발신된 이메일을 랜덤하게 선택하였을 때, 이 중 스팸 메일이 12종 이하일 확률을  $\bar{X}$  의 분포 혹은 근사 분포를 사용하여 계산하시오.

(2) (5점) 이 기관에서 발신된 전체 이메일 중 250을 랜덤하게 선택하였을 때, 이 중 스팸 메일의 비율을  $\hat{p}$  라고 하자.  $\hat{p}$  의 분포를 사용하여 스팸 메일의 비율이 30% 이상일 확률을 구하시오.

6. (10점) 다섯 개의 원소로 구성된 모평균  $\mu=9$ 의 값을 갖는 모집단이 있다. 단순 랜덤 비복원 추출을 사용하여 크기  $n=3$ 의 랜덤 표본을 선택하였을 때, 표본 평균과 표본 중앙값의 확률 분포표가 다음과 같다.

$\bar{X}$	6	7	8	9	10	11	12
$P(\bar{X})$	0.1	0.2	0.2	0.2	0.2	0.1	0.1

$X_{med}$	6	9	12
$P(X_{med})$	0.3	0.4	0.3

(1) (2점) 표본 평균과 표본 중앙값은 모두 모평균에 대한 비편향 추정량임을 보이시오.

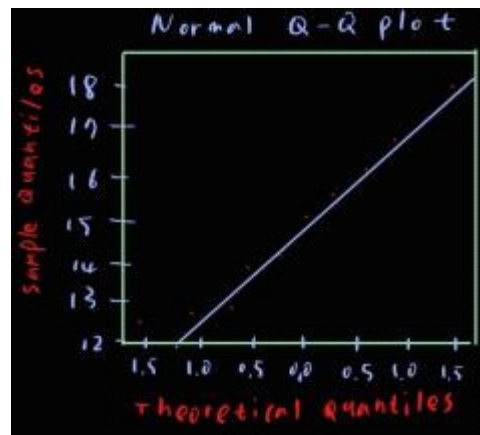
(2) (4점) 표본 평균과 표본 중앙값의 표준오차를 각각 계산하시오.

(3) (4점) 모평균에 대한 두 추정량 중 어느 것이 더 좋은 추정량이라고 생각하는가? 최소분산 불편추정량의 관점에서 그 이유를 간단히 설명하시오.

7. (8점) 한 사회학자는 최근들어 가구원수가 지속적으로 감소하고 있으며, 도시의 평균 가구원수는 2.5명 미만이라고 주장한다. 이를 위해 도시 지역에서 150 가구를 무작위로 추출하여 가구원수를 조사하였고, 그 결과 평균 2.3명, 표준편차 1.5명인 것으로 나타난다. 이 조사결과는 사회학자의 주장을 뒷받침한다고 볼 수 있는가? 유의수준 1%에서 이를 유의확률을 사용하여 검정하시오. 검정 결과의 의미는 무엇인가?

8. (총 13점) 제약 회사는 자신들이 출시한 건강음료가 평균 7.7 ml의 비타민 C 원액을 포함하고 있다고 광고하였다. 하지만 소비자 단체는 이 음료의 비타민 C 원액 함량이 미달이라 생각하고 과대 광고에 대한 조사를 요청하였다. 공정거래위원회는 이를 위해 10개의 제품을 수집하여 조사하였고, 그 결과 표본평균 7.548 ml, 표본 표준편차 0.21 ml 로 나타났다.

(1) (2점) 주어진 표본 자료를 사용하여 가설 검정을 통한 추론을 진행하려고 한다. 이를 위해 필요한 가정사항은 무엇인가? 또한 이러한 가정 사항이 올바르게 적용된다고 볼 수 있겠는가? 오른쪽에 주어진 표본의 정규분포 분위수 대조도를 사용하여 이에 답하시오.



(2) (8점) 소비자 단체의 주장에 대해 올바른 가설을 세우고 유의수준 5%에서 가설검정을 시행하시오. 검정 결과의 의미는 무엇인가?

(3) (3점) 만약 (1)에서 사용한 정규분포 분위수 대조도가 직선의 형태로 나타나지 않았다면, 소비자 단체의 주장에 대한 가설 검정을 시행하기 위해서는 어떤 방법을 쓸 수 있는지 간략히 설명하시오.

9. (8점) 과거 자료에 따르면 A지역에 내린 비의 평균 산성도가 5.2(pH)라고 한다. 최근 들어 이 수치에 변화가 나타났는지를 알아보기 위해 50개의 산성도 조사 자료를 랜덤하게 수집하였고, 그 결과 표본평균이 5.667, 표준편차가 0.921로 나타났다. 이 조사 결과를 통해 최근 이 지역에 내리는 비의 산성도 수치가 변화하였다고 할 수 있는가? 유의수준 5%에서 신뢰구간을 사용하여 이를 검정하시오. 검정 결과의 의미는 무엇인가?

10. (총 11점) 다음 물음에 답하시오.

(1) (2점) 두 확률변수  $X, Y$ 에 대해  $Y=aX+b$ 로 정의하였다.  $\text{Corr}(X, Y)$ 의 값을 구하시오. 이 결과가 의미하는 것은 무엇인가?

(2) (4점) 모평균이 50이고 모분산이 14.5로 알려진 정규 모집단에서 크기  $n=20$ 의 랜덤 표본을 선택하여 모평균이 50이 아니라는 주장에 대해 검정하려고 한다. 이 때, 표본평균이  $48.5 \leq \bar{X} \leq 51.5$  인 경우에는 귀무가설을 기각할 수 없다고 한다. 이 검정의 제 1종 오류를 구하시오.

(3) (5점) 확률변수  $X \sim N(4, 2^2)$ 에 대해  $P(X^2 - 8X < 2)$ 의 값을 구하시오.