

# 통계학개론

문 1. 사건  $A$ 가 일어날 확률은 0.6이고, 사건  $A$ 가 일어났을 때 사건  $B$ 가 일어날 조건부확률은 0.4이다. 사건  $A$ 와  $B$ 가 서로 독립일 때, 사건  $B$ 가 일어날 확률은?

- ① 0.24
- ② 0.4
- ③ 0.5
- ④ 0.6

문 2. 확률변수  $X$ 와  $Y$ 의 분산이 각각 25와 36이고  $X$ 와  $Y$ 의 공분산은  $-20$ 일 때,  $2X+1$ 과  $Y-5$  사이의 상관계수는?

- ①  $-\frac{2}{3}$
- ②  $-\frac{1}{3}$
- ③  $\frac{1}{3}$
- ④  $\frac{2}{3}$

문 3. 평균이  $\mu$ 이고 분산이 4인 정규모집단에서  $\mu$ 에 대한 95% 신뢰구간을 추정하고자 한다. 크기가 16과 64인 임의표본(random sample)으로부터 추정된 신뢰구간의 길이를 각각  $A$ 와  $B$ 라고 할 때,  $\frac{B}{A}$ 의 값은?

- ①  $\frac{1}{4}$
- ②  $\frac{1}{2}$
- ③ 2
- ④ 4

문 4. 절편이 있는 단순선형회귀분석에서 결정계수(coefficient of determination)에 대한 설명으로 옳지 않은 것은?

- ① 항상 0보다 크거나 같고 1보다 작거나 같다.
- ② 종속변수와 독립변수 사이의 표본상관계수 제곱과 같다.
- ③ 종속변수의 분산에 대한 불편추정량(unbiased estimator)이다.
- ④ 총변동(total sum of squares) 중에서 추정된 회귀식에 의해 설명되는 변동(regression sum of squares)의 비율을 나타낸다.

문 5.  $X_1, X_2, \dots, X_n$ 이 어떤 모집단으로부터의 임의표본(random sample)

일 때, 표본평균  $\bar{X} = \frac{1}{n} \sum_{i=1}^n X_i$ 이고 표본분산  $S^2 = \frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X})^2$ 이다. 다음 통계량을 이용하여 계산한 표본분산이  $a$ 이고 변동계수(또는 변이계수)가  $b$ 이면,  $a+b$ 는? (단, 변동계수는 백분율(%)로 환산하지 않은 값으로 한다)

$$n = 3, \quad \bar{X} = 2, \quad \sum_{i=1}^n X_i^2 = 20$$

- ① 3
- ② 4
- ③ 5
- ④ 6

문 6. 분산이 같고 서로 독립인 두 정규모집단  $A$ 와  $B$ 로부터 크기가 각각 16인 표본을 임의로 추출하여 두 모평균의 차에 대한 검정을 하려고 한다. 모집단  $A$ 의 표본분산이 64이고 모집단 공통분산의 추정량인 합동표본분산(pooled sample variance)이 50일 때, 모집단  $B$ 의 표본분산은? (단, 확률변수  $X_1, X_2, \dots, X_n$ 의 표본평균

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

- ① 16
- ② 25
- ③ 36
- ④ 49

문 7. 다음은 단순선형회귀모형  $Y_i = \beta_0 + \beta_1 X_i + \epsilon_i, \quad i = 1, 2, \dots, n$ 을 적합할 때 나타나는 분산분석표의 일부이다. 이에 대한 설명으로 옳지 않은 것은? (단,  $\epsilon_i$ 는 평균이 0이고 분산이  $\sigma^2$ 인 정규분포를 따르며, 서로 독립이다)

요인	제곱합	자유도	평균제곱	F-값	p-값
회귀			20		0.0064
잔차	30	15			

- ① F-값은 10이다.
- ② 표본의 크기  $n$ 은 17이다.
- ③ 종속변수와 독립변수 사이의 표본상관계수는 0.4이다.
- ④ 유의수준 5%에서 '이 회귀모형은 유의하지 않다'는 귀무가설을 기각한다.

문 8. 평균이  $\mu$ 이고 분산이  $\sigma^2$ 인 정규모집단에서의 임의표본(random sample)  $X_1, X_2, \dots, X_n$ 에 대하여 표본평균  $\bar{X} = \frac{1}{n} \sum_{i=1}^n X_i$ 이고 표본분산  $S^2 = \frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X})^2$ 일 때, 다음 중 확률변수의 분포에 대한 설명으로 옳지 않은 것은? (단, 표본의 크기  $n$ 은 1보다 크다)

- ①  $\frac{\sqrt{n}(\bar{X} - \mu)}{S}$ 는 자유도가  $n$ 인  $t$ -분포를 따른다.
- ②  $\frac{\bar{X}}{\sigma}$ 는 평균이  $\frac{\mu}{\sigma}$ 이고 분산이  $\frac{1}{n}$ 인 정규분포를 따른다.
- ③  $\sum_{i=1}^n X_i$ 는 평균이  $n\mu$ 이고 분산이  $n\sigma^2$ 인 정규분포를 따른다.
- ④  $\frac{\sqrt{n}(\bar{X} - \mu)}{\sigma}$ 는 평균이 0이고 분산이 1인 표준정규분포를 따른다.

문 9. 어느 지역에서의 특정 스마트폰 모델에 대한 남녀별 선호도 차이를 알아보기 위해 남자 200명, 여자 100명을 임의로 추출하여 조사한 결과가 다음 표와 같다. 이를 검정하기 위한  $Z$ -검정통계량의 값은? (단,  $Z$ -검정통계량은 근사적으로 표준정규분포를 따른다)

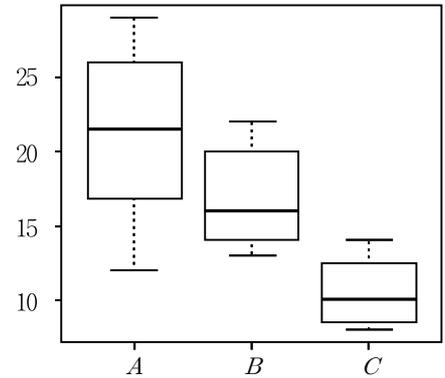
	표본크기	특정 스마트폰 모델 선호자 수
남자	200	100
여자	100	40

- ①  $\frac{|0.5 - 0.4|}{\sqrt{\left(\frac{15}{30} \times \frac{15}{30}\right) \times \frac{1}{140}}}$
- ②  $\frac{|0.5 - 0.4|}{\sqrt{\left(\frac{15}{30} \times \frac{15}{30}\right) \times \frac{1}{300}}}$
- ③  $\frac{|0.5 - 0.4|}{\sqrt{\left(\frac{14}{30} \times \frac{16}{30}\right) \times \left(\frac{1}{200} + \frac{1}{100}\right)}}$
- ④  $\frac{|0.5 - 0.4|}{\sqrt{\frac{14}{30} \times \left(\frac{1}{200} + \frac{1}{100}\right)}}$

문 10. 반복수가 같지 않은 일원배치 분산분석법(one-way analysis of variance)에서  $i$ 번째 처리의  $j$ 번째 자료인  $Y_{ij}$ 의 분산이  $\sigma^2$ 이고  $i$ 번째 처리의 평균  $\bar{Y}_i = \frac{1}{n_i} \sum_{j=1}^{n_i} Y_{ij}$ 이고 총평균  $\bar{Y}_{..} = \frac{1}{n} \sum_{i=1}^k \sum_{j=1}^{n_i} Y_{ij}$ 일 때,  $\sigma^2$ 의 불편추정량(unbiased estimator)은? (단,  $i = 1, 2, \dots, k$  이고,  $j = 1, 2, \dots, n_i$  이고,  $n_i > 1$ 이며, 표본의 크기  $n$ 은  $n_1 + n_2 + \dots + n_k$ 이다)

- ①  $\frac{\sum_{i=1}^k \sum_{j=1}^{n_i} (Y_{ij} - \bar{Y}_i)^2}{n - k}$
- ②  $\frac{\sum_{i=1}^k \sum_{j=1}^{n_i} (Y_{ij} - \bar{Y}_{..})^2}{n - k}$
- ③  $\frac{\sum_{i=1}^k \sum_{j=1}^{n_i} (Y_{ij} - \bar{Y}_i)^2}{n - 1}$
- ④  $\frac{\sum_{i=1}^k \sum_{j=1}^{n_i} (Y_{ij} - \bar{Y}_{..})^2}{n - 1}$

문 11. 어느 회사의 기계 A, B, C로부터 생산된 베어링의 수명에 대한 상자그림(box plot)은 다음과 같다. 이에 대한 설명으로 옳지 않은 것은?



- ① 최댓값은 A가 C보다 크다.
- ② 범위(range)는 A가 B보다 크다.
- ③ 중앙값(median)의 크기는  $A > B > C$  순이다.
- ④ B의 제3사분위수( $Q_3$ )는 A의 제1사분위수( $Q_1$ )보다 작다.

문 12. 두 이산확률변수  $X$ 와  $Y$ 의 결합확률분포표가 다음과 같을 때,  $X$ 와  $Y$ 의 공분산은?

$Y \backslash X$	-2	0	2
0	0	$\frac{1}{3}$	0
1	$\frac{1}{3}$	0	$\frac{1}{3}$

- ① 0
- ②  $\frac{1}{3}$
- ③  $\frac{2}{3}$
- ④ 1

문 13. 어느 불펜 제조 공장에서 100개의 표본을 임의로 추출하여 불량 여부를 조사한 결과 12개가 불량품이었다. 이항분포의 정규분포근사를 이용하여 불량품의 비율에 대한 95% 신뢰구간을 구한 것은? (단,  $Z$ 가 표준정규분포를 따르는 확률변수일 때,  $P(|Z| < 1.645) = 0.90$ 이고  $P(|Z| < 1.96) = 0.95$ 이다)

- ①  $\frac{12}{100} \pm 1.96 \sqrt{\frac{12}{100} \times \left(1 - \frac{12}{100}\right)}$
- ②  $\frac{12}{100} \pm 1.96 \sqrt{\frac{12}{100} \times \left(1 - \frac{12}{100}\right) \times \frac{1}{100}}$
- ③  $\frac{12}{100} \pm 1.645 \sqrt{\frac{12}{100} \times \left(1 - \frac{12}{100}\right)}$
- ④  $\frac{12}{100} \pm 1.645 \sqrt{\frac{12}{100} \times \left(1 - \frac{12}{100}\right) \times \frac{1}{100}}$

문 14.  $X_1, X_2, X_3$ 은 평균이  $\mu$ 인 모집단에서의 임의표본(random sample)이다. 다음의  $\mu$ 에 대한 추정량 중 불편추정량(unbiased estimator)만을 모두 고른 것은?

$$Y_1 = \frac{X_1 + X_2 + X_3}{3}, \quad Y_2 = \frac{2X_1 + X_2}{3}, \quad Y_3 = X_3$$

- ①  $Y_1$
- ②  $Y_1, Y_2$
- ③  $Y_1, Y_3$
- ④  $Y_1, Y_2, Y_3$

문 15. 확률변수  $X$ 는 평균이 4이고 분산이 9인 정규분포를 따르고, 확률변수  $Y$ 는 평균이 2이고 분산이 3인 정규분포를 따른다.  $X$ 와  $Y$ 가 서로 독립일 때, 다음 설명 중 옳은 것은?

- ①  $2X+1$ 의 분산은 18이다.
- ②  $X+2$ 와  $2Y+1$  사이의 상관계수는 0보다 크다.
- ③  $X-Y$ 는 평균이 2이고 분산이 6인 정규분포를 따른다.
- ④  $\frac{X+Y}{2}$ 는 평균이 3이고 분산이 3인 정규분포를 따른다.

문 16.  $X_1, X_2, \dots, X_n$ 은 평균이 4이고 분산이 25인 정규모집단에서의 임의표본(random sample)이다.  $P(\bar{X} \leq 4+5y) = 0.5$ 를 만족하는  $y$ 값은? (단, 표본의 크기  $n$ 은 16이고, 표본평균  $\bar{X} = \frac{1}{n} \sum_{i=1}^n X_i$ 이다)

- ①  $-\frac{4}{5}$
- ② 0
- ③  $\frac{4}{5}$
- ④ 1

문 17. 확률변수  $X$ 는 평균이 3이고 분산이 4이다.  $E\left(\frac{X-3}{2}\right) = p$ 라 하고

$$E\left[\left(\frac{X-3}{2}\right)^2\right] = q$$
라 할 때,  $p+q$ 는?

- ① 0
- ② 1
- ③ 2
- ④ 3

문 18. 카이제곱분포에 대한 설명으로 옳은 것만을 모두 고른 것은?

ㄱ.  $Z$ 가 표준정규분포를 따르는 확률변수일 때,  $Z^2$ 은 자유도가 1인 카이제곱분포를 따른다.

ㄴ. 서로 독립인 두 확률변수가 각각 카이제곱분포를 따를 때, 두 확률변수의 합도 카이제곱분포를 따른다.

ㄷ. 두 범주형자료의 독립성검정에서 행의 수가 3이고 열의 수가 4인 분할표(contingency table)를 이용할 때, 카이제곱검정통계량은 자유도가 12인 카이제곱분포를 따른다.

- ① ㄱ
- ② ㄱ, ㄴ
- ③ ㄴ, ㄷ
- ④ ㄱ, ㄴ, ㄷ

문 19. 다음은 어느 기관에서 세 가지 교육방법에 따른 근무평가점수의 차이를 알아보기로 하자. 반복수가 같은 일원배치 분산분석법(one-way analysis of variance)을 적용하여 얻은 분산분석표의 일부이다. 이에 대한 설명으로 옳지 않은 것은?

요인	제곱합	자유도	평균제곱	F-값	p-값
처리	㉠	㉡			㉢
잔차			㉣		
계	㉤				

- ① ㉡은 2이다.
- ② 근무평가점수의 분산에 대한 추정값은 ㉣이다.
- ③ ㉣이 고정되어 있을 때, ㉠이 커지면 ㉢은 작아진다.
- ④ ㉢이 0.05보다 크면 유의수준 5%에서 세 가지 교육방법에 따른 근무평가점수는 차이가 있다고 할 수 있다.

문 20. 평균이 172이고 분산이 100인 정규모집단에서 10개의 표본을 임의로 추출할 때, 모평균보다 큰 표본의 수를 확률변수  $X$ 라고 하자.  $X$ 의 분산은?

- ① 1
- ②  $\frac{5}{2}$
- ③  $\sqrt{10}$
- ④ 10