

통 계 학

문 1. 어떤 학교에서 35명의 학생을 무작위로 추출하여 수학 성적을 조사한 결과 평균과 중앙값은 각각 70점, 76점이었고, 동점자는 없었다. 5명의 학생을 추가로 뽑아 수학 성적을 조사하여 다음 값을 얻었을 때, 40명 학생의 수학 성적 평균과 중앙값을 바르게 연결한 것은?

77, 76, 61, 81, 75

	평균	중앙값
①	70.5	75
②	70.5	76
③	71.0	75
④	71.0	76

문 2. 다음은 헌혈한 사람 중 80명을 무작위로 뽑아 혈액형을 정리한 표이다. 헌혈한 사람의 혈액형 비(O형, A형, B형, AB형 순)가 3:4:2:1이라는 귀무가설을 유의수준 10%에서 검정하기 위한 검정통계량의 값과 귀무가설의 검정결과를 바르게 연결한 것은? (단, $\chi^2_\alpha(k)$ 는 자유도가 k 인 카이제곱분포의 $(1 - \alpha)$ 분위수를 나타내고, $\chi^2_{0.1}(3) = 6.25$, $\chi^2_{0.1}(4) = 7.78$ 이다)

혈액형	O	A	B	AB	합계
빈도	26	40	8	6	80

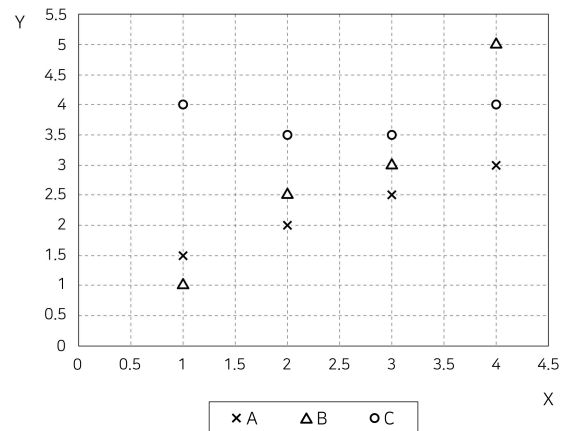
검정통계량의 값	귀무가설의 검정결과
① $\frac{20}{3}$	기각함
② $\frac{20}{3}$	기각하지 못함
③ $\frac{43}{6}$	기각함
④ $\frac{43}{6}$	기각하지 못함

문 3. 자료 (x_i, y_i) ($i = 1, 2, \dots, 20$)로부터 다음 값을 얻었을 때,
이 자료를 최소제곱법으로 적합한 단순선회귀식은?

$$\begin{aligned} \sum_{i=1}^{20} x_i &= 100, & \sum_{i=1}^{20} x_i^2 &= 1,200, & \sum_{i=1}^{20} y_i &= 1,000, \\ \sum_{i=1}^{20} y_i^2 &= 70,000, & \sum_{i=1}^{20} x_i y_i &= 6,400 \end{aligned}$$

- ① $\hat{Y} = 40 + 2x$
- ② $\hat{Y} = 20 + 2x$
- ③ $\hat{Y} = 40 + x$
- ④ $\hat{Y} = 20 + x$

문 4. 다음은 A, B, C 표본의 두 변수 X 와 Y 에 대한 산점도이다. X 와 Y 의 피어슨(Pearson) 표본상관계수가 큰 것부터 순서대로 나열한 것은?



- ① A, B, C
- ② A, C, B
- ③ B, A, C
- ④ C, B, A

문 5. 두 정규모집단 $N(\mu_1, \sigma_1^2)$ 과 $N(\mu_2, \sigma_2^2)$ 으로부터 각각 크기가 n_1 , n_2 이며 독립인 확률표본을 뽑아 가설 $H_0: \sigma_1^2 = \sigma_2^2$ 대 $H_1: \sigma_1^2 > \sigma_2^2$ 을 검정통계량 $F_0 = \frac{S_1^2}{S_2^2}$ 을 사용하여 유의수준 α 로 검정할 때, 귀무가설의 기각역은? (단, S_1^2 과 S_2^2 은 두 표본의 분산이고, $F_\alpha(k_1, k_2)$ 는 자유도가 k_1, k_2 인 F 분포의 $(1 - \alpha)$ 분위수를 나타낸다)

- $$\begin{aligned} \textcircled{1} \quad & F_0 > F_{1-\alpha}(n_1-1, n_2-1) \\ \textcircled{2} \quad & F_0 > F_{\alpha}(n_2-1, n_1-1) \\ \textcircled{3} \quad & F_0 > \frac{1}{F_{\alpha}(n_1-1, n_2-1)} \\ \textcircled{4} \quad & F_0 > \frac{1}{F_{1-\alpha}(n_2-1, n_1-1)} \end{aligned}$$

문 6. 주머니 안에 흰 공과 검은 공이 섞여 있다. 전체 공 10개 중 흰 공의 개수에 대한 다음 가설을 고려한다.

- 귀무가설(H_0): 흰 공의 개수는 5개 이하이다.
- 대립가설(H_1): 흰 공의 개수는 6개 이상이다.

주머니에서 공 세 개를 무작위로 비복원추출한 후, 세 개의 공이 모두 흰색이면 귀무가설을 기각한다. 이때 제2종 오류를 범할 확률의 최댓값은?

- $$\begin{array}{ll} \textcircled{1} \quad \frac{1}{12} & \textcircled{2} \quad \frac{1}{6} \\ \textcircled{3} \quad \frac{5}{6} & \textcircled{4} \quad \frac{11}{12} \end{array}$$

문 7. Y 는 이항분포 $B(n, p)$ 를 따르는 확률변수이다. 성공 확률 p 의

추정량 $\hat{p}$ 을 $\frac{Y+2}{n+3}$ 라고 할 때, 편의(bias)는?

- ① $\frac{-3p}{n+3}$
 ② 0
 ③ $\frac{1-2p}{n+3}$
 ④ $\frac{2-3p}{n+3}$

문 8. $X_1, X_2, \dots, X_{10}$ 은 구간 $(2, 4)$ 에서의 균일분포(uniform distribution)로부터 뽑힌 확률표본이다. 이에 대한 설명으로 옳은 것만을 모두 고르면? (단, $X_{(10)} = \max\{X_1, X_2, \dots, X_{10}\}$ 이다)

$$\neg. E(X_1^2) = \frac{28}{3}$$

$$\neg. 2 < x < 4 \text{ 일 때, } P(X_{10} > x) = 2 - \frac{1}{2}x \text{ 이다.}$$

$$\neg. E(X_{(10)}) = \frac{19}{5}$$

- ① $\neg$, $\neg$
 ② $\neg$, $\neg$
 ③ $\neg$, $\neg$
 ④ $\neg$, $\neg$, $\neg$

문 9. 400명의 사람을 무작위로 뽑아 200명씩 두 그룹으로 나누고 각 그룹에 서로 다른 치료제 A, B를 복용시킨 후 치유된 비율을 조사한 결과, 치료제 A의 치유율은 40%, 치료제 B의 치유율은 60%로 나타났다. p_A, p_B 를 각각 치료제 A, B의 모집단 치유율이라 하고 가설 $H_0: p_A = p_B$ 대 $H_1: p_A < p_B$ 를 검정할 때, p -값(유의확률)은? (단, $\Phi(x)$ 는 표준정규분포의 누적분포함수이다)

- ① $\Phi(-1)$
 ② $\Phi(-2)$
 ③ $\Phi(-3)$
 ④ $\Phi(-4)$

문 10. 100명 중 1명꼴로 감염되는 질병이 있다. 이 질병에 대한 감염 여부 검사에서 질병에 감염된 사람은 90%의 확률로, 질병에 감염되지 않은 사람은 5%의 확률로 양성으로 판정된다. 검사의 정확성을 높이기 위해 같은 검사를 두 번 실시하기로 한다. 무작위로 뽑힌 어떤 사람이 두 번의 검사에서 모두 양성으로 판정되었을 때, 실제로 그 질병에 감염되었을 확률은? (단, 두 번의 검사는 서로 독립이고, 양성 판정 확률은 동일하다고 가정한다)

- ① $\frac{2}{13}$
 ② $\frac{9}{20}$
 ③ $\frac{361}{757}$
 ④ $\frac{36}{47}$

문 11. 다음은 직장인 40명을 무작위로 뽑아 네 가지의 점심 메뉴 중 선호하는 메뉴를 조사한 것이다.

메뉴	인원 수
김치찌개	18
돈가스	10
햄버거	7
자장면	5

이 자료를 바탕으로 “네 가지 메뉴에 대한 선호도에 차이가 없다.”라는 귀무가설을 검정하려고 한다. 관측도수와 귀무가설 하에서의 기대도수 간 차이를 제곱한 값이 가장 큰 메뉴는?

- ① 김치찌개
 ② 돈가스
 ③ 햄버거
 ④ 자장면

문 12. 다음 A, B, C, D 4개 지역에서 5명씩을 무작위로 뽑아 조사한 출근 소요시간에 대한 설명으로 옳지 않은 것은?

지역	출근 소요시간(분)
A	10, 20, 30, 40, 50
B	15, 25, 35, 45, 55
C	20, 40, 60, 80, 100
D	15, 35, 55, 75, 95

- ① A와 B 지역의 표준편차는 같다.
 ② A와 C 지역의 변동계수는 같다.
 ③ B와 C 지역의 변동계수는 같다.
 ④ C와 D 지역의 표준편차는 같다.

문 13. 다음은 자료 $(x_{1i}, x_{2i}, x_{3i}, y_i)$ ($i = 1, 2, \dots, 38$)에

다중선형회귀모형(multiple linear regression model)

$Y_i = \beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + \beta_3 x_{3i} + \epsilon_i$ 를 최소제곱법으로 적합하여 얻은 회귀분석 결과이다. 이에 대한 설명으로 옳지 않은 것은? (단, ϵ_i 는 $N(0, \sigma^2)$ 을 따르고, 서로 독립이다)

요인	제공합	자유도	평균제공	F-값	p-값
회귀	4540.6	3	1513.5	22.8	0.000
잔차	2262.0	34	66.5		
전체	6802.6	37			

모형	비표준화 계수	t-값	p-값
상수	77.89	3.79	0.001
x_1	-19.33	-4.64	0.000
x_2	0.48	7.93	0.000
x_3	-0.11	-0.66	0.514

- ① 종속변수(Y)의 전체 변동 중에서 이 모형에 의하여 설명되는 변동이 차지하는 비율은 $\frac{4540.6}{6802.6}$ 이다.
- ② 유의수준 1%에서 이 모형은 유의하다.
- ③ 유의수준 5%에서 가설 $H_0: \beta_3 = 0$ 대 $H_1: \beta_3 \neq 0$ 을 검정할 때, H_0 를 기각하지 못한다.
- ④ 추정된 회귀식은 $\hat{Y} = 77.89 - 19.33x_1 + 0.48x_2$ 이다.

문 14. 세 가지 다이어트 방법에 따라 체중 감량에 차이가 있는지를 알아보기 위해 15명의 사람을 무작위로 뽑아 5명씩 세 그룹으로 나눈 후 각 그룹에 서로 다른 다이어트 방법을 적용하여 실험을 진행하였다. 자료를 분석하여 다음과 같은 분산분석표를 얻었을 때, 이에 대한 해석으로 옳은 것은? (단, $F_\alpha(k_1, k_2)$ 는 자유도가 k_1, k_2 인 F분포의 $(1 - \alpha)$ 분위수를 나타내고, $F_{0.1}(2, 12) = 2.81$, $F_{0.05}(2, 12) = 3.89$, $F_{0.01}(2, 12) = 6.93$ 이다)

요인	제공합	자유도	평균제공	F-값
처리			150	
오차				
전체	540			

- ① 유의수준 10%에서 차이가 없다.
- ② 유의수준 10%에서는 차이가 있으나 유의수준 5%에서는 차이가 없다.
- ③ 유의수준 5%에서는 차이가 있으나 유의수준 1%에서는 차이가 없다.
- ④ 유의수준 1%에서 차이가 있다.

문 15. 콘크리트 건조방법(A)과 혼합방법(B)이 압축강도에 영향을 미치는지를 알아보려고 한다. 건조방법은 3수준, 혼합방법은 4수준을 고려하고, 각 수준의 조합에서 3회 반복 실험하여 다음과 같은 분산분석표를 얻었을 때, 유의수준 5%에서 유의한 요인만을 모두 나열한 것은? (단, $A \times B$ 는 인자 A, 인자 B의 교호작용(interaction)을 나타내고, $F_\alpha(k_1, k_2)$ 는 자유도가 k_1, k_2 인 F분포의 $(1 - \alpha)$ 분위수를 나타내며, $F_{0.05}(2, 24) = 3.40$, $F_{0.05}(3, 24) = 3.01$, $F_{0.05}(6, 24) = 2.51$ 이다)

요인	제공합	자유도	평균제공	F-값
인자 A	8.0			
인자 B	8.7			
$A \times B$	15.6			
오차	24.0			

- ① 인자 A, 인자 B
- ② 인자 A, $A \times B$
- ③ 인자 B, $A \times B$
- ④ 인자 A, 인자 B, $A \times B$

문 16. 다음은 어느 나라의 2019 ~ 2021년의 1인당 국민소득이다. 이 기간 동안 1인당 국민소득의 연평균 증가율은?

연도	2019	2020	2021
1인당 국민소득(달러)	10,000	15,000	30,000

- ① $100(\sqrt{3} - 1)\%$
- ② $100\sqrt{3}\%$
- ③ 75%
- ④ 175%

- 문 17. A, B 지역에서 각각 “갑, 을, 병, 정 네 후보의 지지도는 동일하다.”라는 귀무가설을 검정하기 위해 A 지역에서 110명, B 지역에서 1,100명을 무작위로 뽑아 조사한 지지자 수가 다음과 같을 때, 이에 대한 설명으로 옳은 것만을 모두 고르면? (단, χ_A^2 , χ_B^2 는 각각 A, B 지역에서의 카이제곱 통계량 값이고, p_A , p_B 는 각각 A, B 지역에서의 p -값(유의확률)이다)

지역 \ 후보	갑	을	병	정
A	30	40	25	15
B	300	400	200	200

ㄱ. 위의 내용은 동질성검정에 해당한다.

ㄴ. $\frac{\chi_B^2}{\chi_A^2} < 10$

ㄷ. $\frac{p_B}{p_A} < 1$

- ① ㄱ, ㄴ
 ② ㄱ, ㄷ
 ③ ㄴ, ㄷ
 ④ ㄱ, ㄴ, ㄷ

- 문 18. 실내온도가 생산량에 영향을 미치는지를 알아보기 위하여 5가지의 실내온도에서 6회씩 실험을 실시하여 총 30개의 측정값을 얻었다. 일원배치 분산분석을 수행한 결과 처리평균제곱의 값이 250이고, 오차제곱합이 처리제곱합의 1.5배일 때, “실내온도는 생산량에 영향을 미치지 않는다.”라는 귀무가설을 검정하기 위한 F -검정통계량의 값은?

- ① $\frac{23}{6}$
 ② $\frac{25}{6}$
 ③ $\frac{23}{5}$
 ④ $\frac{26}{5}$

- 문 19. 확률변수 X 와 Y 의 표준화된 확률변수를 각각 W 와 Z 라 하자. 이에 대한 설명으로 옳은 것만을 모두 고르면?

ㄱ. $Cov(Z, W) = Corr(X, Y)$
 ㄴ. $Corr(W, Z) = Cov(W, Z)$
 ㄷ. $Cov(W, Z) = -1$ 이면 $P(W + Z = 0) = 1$ 이다.

- ① ㄱ, ㄴ
 ② ㄱ, ㄷ
 ③ ㄴ, ㄷ
 ④ ㄱ, ㄴ, ㄷ

- 문 20. 주머니 안에 흰 공 1개, 검은 공 2개가 들어 있다. 공을 한 개씩 무작위로 꺼낼 때 꺼낸 공이 검은 공이면 그 공을 흰 공으로 교체하여 주머니 속에 넣는다. 처음으로 흰 공이 나올 때까지 꺼낸 횟수를 확률변수 X 라고 할 때, $E(9X)$ 의 값은?

- ① 13
 ② 16
 ③ 17
 ④ 20

- 문 21. 자료 (x_i, y_i) ($i = 1, 2, \dots, 7$)에 단순선형회귀모형 $Y_i = a + bx_i + \epsilon_i$ 를 최소제곱법으로 적합한 결과, 회귀제곱합은 5, 오차(잔차)제곱합은 5이었다. $x_i^* = 5 + \frac{x_i}{10}$, $y_i^* = 5 + 10y_i$ 로 변환하여 단순선형회귀모형 $Y_i^* = c + dx_i^* + \epsilon_i^*$ 를 최소제곱법으로 적합할 때, 이에 대한 설명으로 옳은 것만을 모두 고르면? (단, ϵ_i , ϵ_i^* 는 평균 0, 분산이 각각 σ^2 , $(\sigma^*)^2$ 인 정규분포를 따르는 서로 독립인 확률변수이고, $t_\alpha(k)$ 는 자유도가 k 인 t 분포의 $(1 - \alpha)$ 분위수를 나타내며, $t_{0.05}(5) = 2.02$, $t_{0.025}(5) = 2.57$ 이다)

ㄱ. 총제곱합은 100이다.
 ㄴ. 결정계수는 0.5이다.
 ㄷ. 가설 $H_0: d = 0$ 대 $H_1: d \neq 0$ 을 검정할 때, 유의수준 5%에서 귀무가설(H_0)을 기각하지 못한다.

- ① ㄴ
 ② ㄱ, ㄷ
 ③ ㄴ, ㄷ
 ④ ㄱ, ㄴ, ㄷ

문 22. 확률변수 X 의 누적분포함수 $F(x)$ 가 다음과 같을 때, 이에 대한 설명으로 옳은 것만을 모두 고르면?

$$F(x) = \begin{cases} 0, & x \leq 0 \\ x^2, & 0 < x < \frac{1}{2} \\ 1, & x \geq \frac{1}{2} \end{cases}$$

ㄱ. $F(x)$ 는 $x=0, \frac{1}{2}$ 에서 불연속이다.

ㄴ. $P(X=\frac{1}{2})=0$ 이다.

ㄷ. 구간 $0 < x < \frac{1}{2}$ 에서 X 의 확률밀도함수를 $f(x)$ 라 할 때, $f(x)=2x$ 이다.

- ① ㄴ
 ② ㄷ
 ③ ㄴ, ㄷ
 ④ ㄱ, ㄴ, ㄷ

문 23. $X_1, \dots, X_8$ 과 $Y_1, \dots, Y_8$ 은 각각 두 정규모집단 $N(2, \sigma^2)$ 과 $N(1, \sigma^2)$ 에서 추출한 서로 독립인 확률표본이고, $\bar{X}, \bar{Y}, S_p$ 가 다음과 같을 때, $P(\bar{X}-\bar{Y} \geq 1+c \times S_p)=0.1$ 을 만족하는 상수 c 의 값은? (단, $t_\alpha(k)$ 는 자유도가 k 인 t 분포의 $(1-\alpha)$ 분위수이다)

$$\begin{aligned} \circ \bar{X} &= \frac{1}{8} \sum_{i=1}^8 X_i \\ \circ \bar{Y} &= \frac{1}{8} \sum_{i=1}^8 Y_i \\ \circ S_p &= \sqrt{\frac{1}{14} \left(\sum_{i=1}^8 (X_i - \bar{X})^2 + \sum_{i=1}^8 (Y_i - \bar{Y})^2 \right)} \end{aligned}$$

- ① $\frac{1}{4}t_{0.1}(14)$
 ② $\frac{1}{3}t_{0.1}(14)$
 ③ $\frac{1}{2}t_{0.1}(14)$
 ④ $t_{0.1}(14)$

문 24. 어떤 식품이 혈중 콜레스테롤의 양에 영향을 주는지 알아보기 위해 10명의 피실험자를 무작위로 추출하여 각 피실험자에 대해 식품을 섭취한 경우와 섭취하지 않은 경우의 혈중 콜레스테롤의 양을 측정하였다. “식품을 섭취한 경우와 섭취하지 않은 경우의 혈중 콜레스테롤의 양에 차이가 있다.”라는 대립가설을 검정하기 위한 t -검정통계량의 값이 -1.9 일 때, 이에 대한 설명으로 옳은 것은? (단, $t_\alpha(k)$ 는 자유도가 k 인 t 분포의 $(1-\alpha)$ 분위수를 나타내고, $t_{0.05}(9)=1.83$, $t_{0.025}(9)=2.26$, $t_{0.05}(18)=1.73$, $t_{0.025}(18)=2.10$ 이다)

- ① 대응(쌍체)비교를 실시하며 유의수준 5%에서 차이가 없다.
 ② 대응(쌍체)비교를 실시하며 유의수준 5%에서 차이가 있다.
 ③ 독립인 표본에 의한 두 집단 비교를 실시하며 유의수준 5%에서 차이가 없다.
 ④ 독립인 표본에 의한 두 집단 비교를 실시하며 유의수준 5%에서 차이가 있다.

문 25. 사건 A, B, C 는 서로 독립이고, $P(A \cap B \cap C) = \frac{4}{27}$ 이다.

$P(A)=P(B)=2P(C)$ 일 때 $P(A \cup B) - P(B \cap C)$ 의 값은?

- ① $\frac{2}{9}$
 ② $\frac{2}{3}$
 ③ $\frac{7}{9}$
 ④ $\frac{5}{6}$