

통 계 학

1. 다음은 중학생 100명을 대상으로 하루 평균 게임 시간을 조사하여 작성한 도수분포표의 일부이다. 이에 대한 설명으로 옳은 것은?

하루 평균 게임 시간	도수	상대도수
1시간 미만	25	
1시간 이상 ~ 2시간 미만	35	
2시간 이상 ~ 3시간 미만	(가)	0.21
3시간 이상 ~ 4시간 미만	7	
4시간 이상 ~ 5시간 미만	7	
5시간 이상	(나)	
합계	100	1.00

- ① (가)의 값은 20이다.
- ② (나)의 값은 3이다.
- ③ 하루 평균 게임 시간이 2시간 미만인 학생들의 비율은 0.5보다 작다.
- ④ 하루 평균 1시간 이상 2시간 미만 게임을 하는 학생들의 비율이 가장 높다.

2. X_1 , X_2 , X_3 은 모평균이 μ 인 모집단으로부터 추출한 확률표본일 때, μ 의 불편추정량으로 옳은 것만을 모두 고르면? (단, μ 는 0이 아니다)

ㄱ. $\frac{1}{3}X_1 + \frac{2}{3}X_2 - X_3$
ㄴ. $\frac{1}{3}X_1 - X_2 + \frac{2}{3}X_3$
ㄷ. $X_1 + \frac{1}{2}X_2 - \frac{1}{2}X_3$
ㄹ. $\frac{1}{2}X_1 + \frac{2}{3}X_2 - \frac{1}{6}X_3$

- ① ㄱ, ㄴ
- ② ㄱ, ㄷ
- ③ ㄴ, ㄹ
- ④ ㄷ, ㄹ

3. 다음은 n 개의 자료 (x_i, y_i) 에 단순선형회귀모형 $y_i = \beta_0 + \beta_1 x_i + \epsilon_i$ 를 최소제곱법으로 적합하여 얻은 분산분석표의 일부이다.

요인	제곱합	자유도	평균제곱	F -값
회귀	SSR		MSR	F_0
잔차	SSE	10	MSE	

$\sum_{i=1}^n (x_i - \bar{x})^2 = 9$, $\sum_{i=1}^n (y_i - \bar{y})^2 = 12$, $\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y}) = 6$ 일 때, 옳은 것만을 모두 고르면? (단, ϵ_i 는 $N(0, \sigma^2)$ 을 따르고 서로 독립이다)

ㄱ. $SSR = 4$
ㄴ. $MSE = 0.8$
ㄷ. $F_0 = 5$

- ① ㄱ, ㄴ
- ② ㄱ, ㄷ
- ③ ㄴ, ㄷ
- ④ ㄱ, ㄴ, ㄷ

4. 다음은 4가지 광고유형에 따라 상품판매량의 차이가 있는지를 알아보기 위해 얻은 분산분석표의 일부이다. 이에 대한 설명으로 옳지 않은 것은?

요인	제곱합	자유도	평균제곱	F -값	p -값
처리	36				0.022
오차		20			
계	96				

- ① 처리평균제곱은 12이다.
- ② F -검정통계량의 값은 4이다.
- ③ 총자료의 개수는 23이다.
- ④ 유의수준 5 %에서 광고유형에 따라 상품판매량은 차이가 있다.

5. 어느 대학교 도서관의 작년 자료에 의하면 학부생 중 도서 대출 비율은 1학년 20 %, 2학년 30 %, 3학년 40 %, 4학년 10 %라고 한다. 올해 이 도서관에서 대출한 학부생 1,000명을 임의추출하여 학년별로 대출한 학생 수를 조사한 결과가 다음과 같다. 올해 이 도서관에서 도서를 대출한 학부생들의 학년 비율이 작년과 차이가 있는지를 검정하기 위한 카이제곱 검정통계량의 값은?

학년	1학년	2학년	3학년	4학년	계
학생 수	210	270	440	80	1,000

- ① 10.1
- ② 11.5
- ③ 12.4
- ④ 15.0

6. 기댓값(평균)이 $\frac{13}{8}$ 인 이산형 확률변수 X 의 확률분포가 다음과 같을 때, $\frac{b}{a}$ 의 값은?

x	0	1	2	3	계
$P(X=x)$	$\frac{1}{4}$	a	b	$\frac{1}{4}$	1

- ① 1
- ② 2
- ③ 3
- ④ 4

7. 연속형 확률변수 X 의 확률밀도함수가 다음과 같을 때, $P(X > 4 \mid X > 2)$ 의 값은?

$$f_X(x) = \begin{cases} \frac{1}{4}e^{-\frac{x}{4}}, & x > 0 \\ 0, & \text{그 외} \end{cases}$$

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

8. 두 확률변수 X 와 Y 는 서로 독립이며 각각 성공의 확률이 $\frac{1}{2}$ 인 베르누이(Bernoulli) 분포를 따른다고 할 때, 확률변수 $W = X + Y$ 의 확률질량함수는?

- ① $P(W=k) = \frac{2!}{k!} \left(\frac{1}{2}\right)^k, k=0, 1, 2$
- ② $P(W=k) = \frac{2!}{k!} \left(\frac{1}{2}\right)^2, k=0, 1, 2$
- ③ $P(W=k) = \frac{2!}{k!(2-k)!} \left(\frac{1}{2}\right)^k, k=0, 1, 2$
- ④ $P(W=k) = \frac{2!}{k!(2-k)!} \left(\frac{1}{2}\right)^2, k=0, 1, 2$

9. 10개의 자료 (x_i, y_i) 에 단순선형회귀모형 $y_i = \beta_0 + \beta_1 x_i + \epsilon_i$ 를 최소 제곱법으로 적합할 때, 이에 대한 설명으로 옳은 것만을 모두 고르면? (단, ϵ_i 는 $N(0, \sigma^2)$ 을 따르고 서로 독립이다)

- ㄱ. 설명변수와 반응변수의 표본평균을 각각 $\bar{x}$ 와 $\bar{y}$ 라 할 때, 회귀직선은 항상 $(\bar{x}, \bar{y})$ 를 지난다.
 - ㄴ. 0이 아닌 총제곱합(SST)이 일정할 때, 잔차제곱합(SSE)이 작아지면 결정계수도 작아진다.
 - ㄷ. 가설 $H_0 : \beta_1 = 0$ 대 $H_1 : \beta_1 \neq 0$ 을 검정하기 위한 F 분포의 분자와 분모의 자유도는 각각 1과 8이다.

- ① ㄱ, ㄴ
- ② ㄱ, ㄷ
- ③ ㄴ, ㄷ
- ④ ㄱ, ㄴ, ㄷ

10. 모평균과 모분산이 알려지지 않은 정규모집단으로부터 임의추출한 21개의 표본에 대한 표본분산이 16이다. 모분산이 40보다 작은지를 유의수준 5 %에서 검정하고자 할 때, 카이제곱 검정통계량의 값과 검정 결과를 바르게 연결한 것은? (단, $\chi^2_{\alpha}(k)$ 는 자유도가 k 인 카이제곱 분포의 제 $100 \times (1 - \alpha)$ 백분위수를 나타내고, $\chi^2_{0.05}(20) = 31.41$, $\chi^2_{0.95}(20) = 10.85$, $\chi^2_{0.05}(21) = 32.67$, $\chi^2_{0.95}(21) = 11.59$ 이다)

	카이제곱 검정통계량의 값	검정 결과
①	8	모분산이 40보다 작다고 할 수 없다
②	8	모분산이 40보다 작다고 할 수 있다
③	11	모분산이 40보다 작다고 할 수 없다
④	11	모분산이 40보다 작다고 할 수 있다

11. 자료의 개수가 100인 어느 자료에서 최댓값 a 가 $a + 10$ 으로 되었을 때, 값이 변하지 않는 측도만을 모두 고르면? (단, 자료의 값은 모두 다르다)

- ㄱ. 평균
ㄴ. 중앙값
ㄷ. 분산
ㄹ. 사분위수범위

- ① ㄱ, ㄴ
② ㄱ, ㄷ
③ ㄴ, ㄹ
④ ㄷ, ㄹ

12. 두 확률변수 X 와 Y 의 결합확률분포가 다음과 같다.

$X \backslash Y$	-1	0	1
-1	0	0.4	0
1	0.3	0	0.3

이에 대한 설명으로 옳은 것만을 모두 고르면?

- ㄱ. $E(XY)$ 는 $E(X) \times E(Y)$ 와 같다.
ㄴ. X 와 Y 는 독립이다.
ㄷ. X 와 Y 의 상관계수는 0이다.

- ① ㄱ, ㄴ
② ㄱ, ㄷ
③ ㄴ, ㄷ
④ ㄱ, ㄴ, ㄷ

13. 자유도가 2보다 큰 t 분포에 대한 설명으로 옳은 것만을 모두 고르면?

- ㄱ. 표준정규분포보다 t 분포의 분산이 더 크다.
ㄴ. t 분포는 자유도가 커질수록 표준정규분포에 수렴한다.
ㄷ. t 분포의 확률밀도함수는 0을 중심으로 좌우 대칭이고 종 모양이다.

- ① ㄱ, ㄴ
② ㄱ, ㄷ
③ ㄴ, ㄷ
④ ㄱ, ㄴ, ㄷ

14. 어느 회사의 연마제 품질이 인자 A 의 3가지 수준과 인자 B 의 6가지 수준에 따라 차이가 있는지를 알아보려고 한다. 다음은 인자 A 와 B 에 대하여 반복이 없는 이원배치법으로부터 얻은 분산분석표의 일부이다. 이에 대한 설명으로 옳은 것은? (단, $F_{\alpha}(k_1, k_2)$ 는 분자와 분모의 자유도가 각각 k_1, k_2 인 F 분포의 제 $100 \times (1 - \alpha)$ 백분위수를 나타낼 때, $F_{0.05}(2, 10) = 4.10$, $F_{0.05}(5, 10) = 3.33$, $F_{0.05}(3, 18) = 3.16$, $F_{0.05}(6, 18) = 2.66$ 이다)

요인	제곱합	자유도	평균제곱	F -값
인자 A	14			
인자 B	28			
오차	20			

- ① 평균제곱오차(MSE)는 2.0이다.
② 오차제곱합(SSE)의 자유도는 총실험횟수에서 1을 뺀 값이다.
③ 유의수준 5 %에서 인자 A 의 수준에 따라 연마제 품질에 차이가 있다.
④ 유의수준 5 %에서 인자 B 의 수준에 따라 연마제 품질에 차이가 있다.

15. 연속형 확률변수 X 의 누적분포함수가 다음과 같을 때, X 의 분산은?

$$F_X(x) = \begin{cases} 0, & x \leq 0 \\ x^2, & 0 < x \leq 1 \\ 1, & x > 1 \end{cases}$$

- ① $\frac{1}{18}$
② $\frac{4}{9}$
③ $\frac{1}{2}$
④ $\frac{2}{3}$

16. 다음은 반복이 3회 있는 이원배치법으로부터 얻은 분산분석표의 일부이다. 인자 A의 수준은 4이고 인자 B의 수준은 5일 때, 옳은 것은?

요인	제곱합	자유도	평균제곱	F-값
인자 A	18	(가)		(라)
인자 B	48			(마)
교호작용 $A \times B$	36	(나)		
오차	80		(다)	

- ① (가)의 값과 (라)의 값은 같다.
- ② (나)의 값과 (마)의 값은 같다.
- ③ (다)의 값과 (라)의 값은 같다.
- ④ (라)의 값과 (마)의 값은 같다.

17. 다음은 모분산이 알려지지 않고 서로 독립인 두 정규모집단 A와 B로부터 임의추출한 표본을 조사한 결과이다. 두 모집단 A와 B의 평균 차이인 $\mu_A - \mu_B$ 에 대한 95 % 신뢰구간은? (단, $t_{\alpha}(k)$ 는 자유도가 k인 t 분포의 제 $100 \times (1 - \alpha)$ 백분위수를 나타내고, $t_{0.025}(8) = 2.306$, $t_{0.025}(9) = 2.262$ 이다)

모집단 (모평균, 모분산)	표본의 크기	표본평균	표본표준편차
A (μ_A, σ^2)	4	22	3
B (μ_B, σ^2)	6	18	1

- ① $4 \pm 2.262 \times 4 \sqrt{\frac{5}{12}}$
- ② $4 \pm 2.262 \times 2 \sqrt{\frac{5}{12}}$
- ③ $4 \pm 2.306 \times 4 \sqrt{\frac{5}{12}}$
- ④ $4 \pm 2.306 \times 2 \sqrt{\frac{5}{12}}$

18. n개의 자료 (x_i, y_i) 에 단순선형회귀모형을 적합하여 얻은 추정식이 $\hat{y}_i = 4 + 6x_i$ 일 때, 관측값 $(x_1, y_1) = (3, 24)$ 에 대한 잔차의 제곱값은?

- ① 1
- ② 2
- ③ 4
- ④ 9

19. 절편이 있는 단순선형회귀모형을 최소제곱법으로 적합할 때, 결정계수에 대한 설명으로 옳은 것만을 모두 고르면? (단, 총제곱합(SST)은 0이 아니다)

- ㄱ. 결정계수의 값은 0보다 크거나 같고 1보다 작거나 같다.
ㄴ. 결정계수의 값이 1이면 회귀제곱합(SSR)이 0이다.
ㄷ. 결정계수는 설명변수와 반응변수의 표본상관계수와 동일하다.

- ① ㄱ
- ② ㄱ, ㄴ
- ③ ㄴ, ㄷ
- ④ ㄱ, ㄴ, ㄷ

20. 어느 우체국에 한 시간 동안 들어가는 사람의 수 Y는 평균이 λ인 포아송 분포를 따르는 확률변수이다. 가설 $H_0 : \lambda = 3$ 대 $H_1 : \lambda < 3$ 에 대한 검정에서 기각역으로 ‘ $Y \leq 1$ ’을 사용할 때, 제1종 오류의 확률은?

- ① $2e^{-3}$
- ② $3e^{-3}$
- ③ $4e^{-3}$
- ④ $5e^{-3}$

21. 사교육이 A 대학 합격 여부와 관계가 있는지를 알아보기 위해 A 대학 지원자 300명을 대상으로 설문 조사를 실시하여 다음과 같은 분할표를 얻었다. 사교육을 받은 지원자 중에서 A 대학에 합격할 오즈(odds) 값은?

A 대학 합격 여부 사교육	합격	불합격
받음	83	131
받지 않음	7	79

- ① $\frac{83}{7}$
- ② $\frac{83}{90}$
- ③ $\frac{83}{131}$
- ④ $\frac{83}{214}$

22. 연속형 확률변수 X 의 확률밀도함수가 다음과 같을 때, X 의 중앙값은? (단, $\log$ 는 자연로그함수이다)

$$f_X(x) = \begin{cases} e^{-x}, & x > 0 \\ 0, & \text{그 외} \end{cases}$$

- ① $\log\left(\frac{1}{2}\right)$
- ② $\log(2)$
- ③ $e^{\frac{1}{2}}$
- ④ e^2

23. 두 사람 A와 B가 번갈아 가면서 공정한 주사위를 던져 처음으로 6의 눈이 나온 사람이 이기는 게임을 한다. 첫 번째로 A가 던질 때, B가 이길 확률은?

- ① $\frac{4}{9}$
- ② $\frac{5}{11}$
- ③ $\frac{1}{2}$
- ④ $\frac{5}{9}$

24. 서로 독립인 추정량 $\hat{\theta}_1$ 과 $\hat{\theta}_2$ 은 각각 모수 θ 의 불편추정량이고 $Var(\hat{\theta}_1) = 3Var(\hat{\theta}_2)$ 을 만족한다. 추정량 $\hat{\theta}_3 = a\hat{\theta}_1 + b\hat{\theta}_2$ 이 θ 의 불편 추정량일 때, $\hat{\theta}_3$ 의 분산을 최소화하기 위한 상수 a 와 b 의 값을 바르게 연결한 것은? (단, θ 는 0이 아니고 $Var(\hat{\theta}_1)$ 과 $Var(\hat{\theta}_2)$ 은 각각 0보다 크다)

- | | a | b |
|---|---------------|---------------|
| ① | $\frac{1}{4}$ | $\frac{3}{4}$ |
| ② | $\frac{1}{3}$ | $\frac{2}{3}$ |
| ③ | $\frac{2}{3}$ | $\frac{1}{3}$ |
| ④ | $\frac{3}{4}$ | $\frac{1}{4}$ |

25. 체비셰프 부등식(Chebyshev's inequality) 정리는 ‘확률변수 X 의 평균이 μ 이고 분산이 $\sigma^2(<\infty)$ 일 때, 모든 $k \geq 1$ 에 대하여 $P(|X-\mu| \geq k\sigma) \leq \frac{1}{k^2}$ 이 성립한다’이다. 이산형 확률변수 X 의 확률분포가 다음과 같을 때, 체비셰프 부등식 정리에 의한 $P(|X| \geq 1)$ 의 상한값은?

x	-1	0	1	계
$P(X=x)$	$\frac{1}{8}$	$\frac{6}{8}$	$\frac{1}{8}$	1

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