

통 계 학

1. 다음은 남학생과 여학생 각 15명의 통계학 성적을 줄기-잎 그림으로 나타낸 것이다. 이에 대한 설명으로 옳은 것만을 모두 고르면?

남학생	여학생
2 1	5 4 5
5 4 3	6 6 7 8
8 4 3 2 1	7 4 5 6 7
2 1	8 1 4 5
5 4 3	9 6 7 8

- ㄱ. 중앙값은 남학생이 여학생보다 3점 작다.
 ㄴ. 사분위수범위는 남학생과 여학생이 같다.
 ㄷ. 표준편차는 여학생이 남학생보다 크다.

- ① ㄱ
 ② ㄱ, ㄴ
 ③ ㄴ, ㄷ
 ④ ㄱ, ㄴ, ㄷ
2. 모집단 $\{2, 4\}$ 로부터 임의추출한 확률변수 X 와 모집단 $\{1, 5\}$ 로부터 임의추출한 확률변수 Y 에 대해 $X+Y$ 의 분산은? (단, X 와 Y 는 독립이다)
- ① 4
 ② 5
 ③ 8
 ④ 10
3. 정규분포를 따르는 확률변수 X, Y, W 의 평균과 분산은 다음과 같다.

확률변수	평균	분산
X	2	4
Y	0	9
W	8	16

다음 확률값을 작은 것부터 순서대로 바르게 나열한 것은?

- ㄱ. $P(X \geq 1)$
 ㄴ. $P(Y \leq -2)$
 ㄷ. $P(W \leq 12)$

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

4. 사건 A, B, C 는 서로 독립이고 $P(A) = 0.1$, $P(B) = 0.2$, $P(C) = 0.3$ 일 때, $P(A \cup B \cup C)$ 의 값은?

- ① $\frac{61}{125}$
 ② $\frac{62}{125}$
 ③ $\frac{63}{125}$
 ④ $\frac{64}{125}$

5. 다음은 벼의 품종(A)과 비료(B)가 쌀 생산량에 영향을 미치는지를 알아보기 위하여 벼의 품종은 a 개의 수준, 비료는 b 개의 수준에서 반복이 없는 이원배치법을 적용하여 얻은 분산분석표의 일부이다. a 와 b 의 값을 바르게 연결한 것은?

요인	제곱합	자유도	평균제곱	F -값
인자 A	20		4	2
인자 B				3
오차				
전체	180			

- | | a | b |
|---|-----|-----|
| ① | 5 | 10 |
| ② | 5 | 11 |
| ③ | 6 | 10 |
| ④ | 6 | 11 |

6. 통계부서에 근무하는 직원 9명의 지난 한 해 동안 출장 일수가 다음과 같이 조사되었다.

10, 20, 30, 40, 60, 70, 70, 70, 80

이 관측값을 $x_1, x_2, \dots, x_9$ 라고 할 때, $\sum_{i=1}^9 (x_i - a)^2$ 을 최소로 하는

a 의 값과 $\sum_{i=1}^9 |x_i - b|$ 를 최소로 하는 b 의 값을 바르게 연결한 것은?

- | | a | b |
|---|-----|-----|
| ① | 50 | 60 |
| ② | 50 | 70 |
| ③ | 60 | 60 |
| ④ | 60 | 70 |

7. 다음 자료에 단순선형회귀모형 $Y_i = \beta_0 + \beta_1 x_i + \epsilon_i (i = 1, 2, 3, 4, 5)$ 를 최소제곱법으로 적합하여 얻은 β_0 와 β_1 의 추정값 a 와 b 의 값을 바르게 연결한 것은? (단, ϵ_i 는 $N(0, \sigma^2)$ 을 따르고 서로 독립이다)

X	1	2	3	4	5
Y	2	3	5	6	9

- $\underline{a}$

$\underline{b}$
- ①

$\frac{1}{10}$

$-\frac{17}{10}$
- ②

$\frac{1}{10}$

$\frac{17}{10}$
- ③

$-\frac{1}{10}$

$-\frac{17}{10}$
- ④

$-\frac{1}{10}$

$\frac{17}{10}$

8. 검은 공이 6개, 흰 공이 2개 들어 있는 상자에서 갑과 을이 번갈아가며 임의로 하나씩 공을 꺼낼 때 흰 공을 먼저 꺼낸 사람이 이기는 게임을 한다. 이때 꺼낸 공을 다시 집어넣고 게임을 하는 방법(게임 A)과 꺼낸 공을 다시 집어넣지 않고 게임을 하는 방법(게임 B)을 고려할 수 있다. 첫 번째 공을 갑이 꺼낸다고 했을 때 게임 A에서 갑이 이길 확률(P_A)과 게임 B에서 갑이 이길 확률(P_B)의 차 $P_A - P_B$ 의 값은?

- ① $-\frac{1}{7}$
- ② 0
- ③ $\frac{1}{7}$
- ④ $\frac{2}{7}$

9. $Z_1, Z_2, \dots, Z_6$ 이 표준정규분포로부터의 임의표본일 때, 다음 식을 만족하는 상수 c 의 값은? (단, $F_\alpha(k_1, k_2)$ 는 분자와 분모의 자유도가 각각 k_1 과 k_2 인 F 분포의 제 $100 \times (1 - \alpha)$ 백분위수를 나타낸다)

$$P\left(\frac{Z_1^2 + Z_2^2}{Z_3^2 + Z_4^2 + Z_5^2 + Z_6^2} < c\right) = 0.05$$

- ① $\frac{1}{2} \times \frac{1}{F_{0.05}(4, 2)}$
- ② $\frac{1}{2} \times F_{0.05}(2, 4)$
- ③ $2 \times \frac{1}{F_{0.05}(4, 2)}$
- ④ $2 \times F_{0.05}(2, 4)$

10. 연속형 확률변수 X 와 Y 의 결합확률밀도함수가 다음과 같을 때 $P\left(Y < \frac{1}{2} \mid X > \frac{1}{2}\right)$ 의 값은?

$$f_{X,Y}(x,y) = \begin{cases} 12xy(1-x), & 0 < x < 1, \ 0 < y < 1 \\ 0, & \text{그 외} \end{cases}$$

- ① $\frac{1}{8}$
- ② $\frac{1}{4}$
- ③ $\frac{1}{2}$
- ④ 1

11. 혈압약 A와 B의 효과를 비교하기 위해 고혈압 환자에서 임의추출한 100명 중 임의로 50명을 선택하여 혈압약 A를 투여하고, 나머지 50명에게는 혈압약 B를 투여한 결과는 다음과 같다.

혈압약 \ 효과	있음	없음
A	37	13
B	27	23

“두 혈압약의 효과는 차이가 없다.”는 가정하에서 혈압약 A를 투여한 후 효과가 있는 고혈압 환자수의 기대도수는?

- ① 30
- ② 32
- ③ 34
- ④ 36

12. $X_1, X_2, \dots, X_n$ 은 구간이 $(2,4)$ 인 균일분포(uniform distribution)

로부터의 임의표본이다. n 이 무한히 커질 때 통계량 $\hat{\theta}_n = \frac{n}{\sum_{i=1}^n X_i}$ 이

확률적으로 수렴하는 값은?

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

13. 나무 30 그루에서 측정한 나무둘레(Y), 나무높이(x_1), 나무부피(x_2)의 자료에 다중선형회귀모형 $Y_i = \beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + \epsilon_i$ 를 적합한 결과의 일부는 다음과 같다. 이에 대한 설명으로 옳은 것만을 모두 고르면? (단, ϵ_i 는 $N(0, \sigma^2)$ 을 따르고 서로 독립이다)

Coefficients:

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	10.53105	1.87281	5.623	5.74e-06
x_1	-0.04515	0.02675	-1.688	0.103
x_2	0.20578	0.01157	17.781	< 2e-16

Residual standard error: 0.7481 on () degrees of freedom
Multiple R-squared: 0.9369, Adjusted R-squared: 0.9323
F-statistic: 200.5 on () and () DF, p-value: < 2.2e-16

- ㄱ. 유의수준 5 %에서 나무부피(x_2)가 주어졌을 때, 나무둘레(Y)에 대해 나무높이(x_1)는 유의한 설명변수이다.
ㄴ. 유의수준 5 %에서 나무높이(x_1)가 주어졌을 때, 나무부피(x_2)가 증가할수록 나무둘레(Y)는 증가한다.
ㄷ. 잔차제곱합의 자유도는 29이다.

- ① ㄴ
② ㄱ, ㄷ
③ ㄴ, ㄷ
④ ㄱ, ㄴ, ㄷ

14. 절삭 방법 A, B, C 가 가공 속도에 영향을 미치는지 알아보기 위해 절삭 방법 A, B, C 에서 각각 6, 7, 8 회 실험한 결과를 분석하여 얻은 분산분석표의 일부는 다음과 같다. 이에 대한 설명으로 옳지 않은 것은?

요인	제곱합	자유도	F -값	p -값
처리		(가)	(나)	(다)
오차	36	(라)		
전체	40			

- ① (가)의 값은 2이다.
② (나)의 값은 1이다.
③ (라)의 값은 18이다.
④ (나)의 값이 커지면 (다)의 값도 커진다.

15. 유의수준과 p -값(유의확률)에 대한 설명으로 옳지 않은 것은?

- ① 유의수준 α 의 검정에서 제1종 오류를 범할 확률이 α 보다 작거나 같다.
② p -값은 검정통계량의 관측값을 기준으로 제2종 오류를 범할 확률이다.
③ p -값이 주어진 유의수준보다 작으면 귀무가설을 기각한다.
④ 같은 자료를 사용하여 동일한 가설검정을 하더라도 유의수준의 값에 따라 검정결과가 달라질 수 있다.

16. 변수 X 와 Y 에 대한 자료 (x_i, y_i) ($i = 1, 2, \dots, 6$)에서 X 와 Y 의 표본평균은 각각 3과 4이고, X 와 Y 의 표본표준편차는 1로 동일하다. 이 자료에 단순선형회귀모형 $Y_i = \beta_0 + \beta_1 x_i + \epsilon_i$ 를 적합하여 얻은 β_1 의 최소제곱추정값이 0.5일 때, 이에 대한 설명으로 옳은 것만을 모두 고르면? (단, ϵ_i 는 $N(0, \sigma^2)$ 을 따르고 서로 독립이다)

- ㄱ. β_0 의 최소제곱추정값은 2.5이다.
ㄴ. 잔차제곱합의 자유도는 4이다.
ㄷ. 결정계수(R^2)는 0.25이다.

- ① ㄱ
② ㄱ, ㄴ
③ ㄴ, ㄷ
④ ㄱ, ㄴ, ㄷ

17. 지역 A, B, C 에서 특정 정책에 대한 지지 여부를 조사한 결과가 다음과 같을 때, 지역과 지지 여부의 독립성을 검정하기 위한 카이제곱 검정통계량의 값과 유의수준 5 %에서의 검정결과를 바르게 연결한 것은? (단, $\chi^2_{\alpha}(k)$ 는 자유도가 k 인 카이제곱 분포의 제 $100 \times (1 - \alpha)$ 백분위수를 나타내고, $\chi^2_{0.05}(1) = 3.84$, $\chi^2_{0.05}(2) = 5.99$, $\chi^2_{0.05}(6) = 12.59$ 이다)

지역 \ 지지 여부	지지함	지지하지 않음
A	30	20
B	40	10
C	50	50

- | | 카이제곱 검정통계량의 값 | 검정결과 |
|---|---------------|---------------------|
| ① | 12.5 | 지역과 지지 여부는 독립이다. |
| ② | 12.5 | 지역과 지지 여부는 독립이 아니다. |
| ③ | 16.5 | 지역과 지지 여부는 독립이다. |
| ④ | 16.5 | 지역과 지지 여부는 독립이 아니다. |

18. 확률변수 X 에 대하여 $E(X) = \frac{1}{p}$ 이고 $Var(X) = \frac{1-p}{p^2}$ 일 때,

$Var(X)$ 의 불편추정량은?

- ① $\frac{X(X-1)}{2}$
- ② $\frac{X^2-1}{2}$
- ③ $\frac{X^2}{2}$
- ④ $\frac{X}{2}$

19. 평균이 μ 이고 분산은 $\sigma^2 = 144$ 인 모집단으로부터 64개의 표본을 임의추출하여 가설 $H_0 : \mu = 30$ 대 $H_1 : \mu > 30$ 를 유의수준 5 %에서 검정하고자 한다. $\mu = 39$ 에서 제2종 오류를 범할 확률은? (단, z_{α} 는 표준정규분포의 제 $100 \times (1 - \alpha)$ 백분위수이다)

- ① $P(Z > z_{0.05} + 4)$
- ② $P(Z < z_{0.05} - 4)$
- ③ $P(Z > z_{0.05} + 6)$
- ④ $P(Z < z_{0.05} - 6)$

20. n 개의 자료 (x_i, y_i) 에 단순선형회귀모형 $Y_i = \beta_0 + \beta_1 x_i + \epsilon_i$ 를 최소제곱법으로 적합하여 얻은 회귀식은 $\hat{y}_i = 2 + 3x_i$ 이었다. 이 회귀식에 대한 회귀제곱합은 640이고 잔차제곱합은 360이었다. 자료 (x_i, y_i) 를 $x_i^* = 1 + 3x_i$, $y_i^* = 5 - \frac{y_i}{10}$ 로 변환하여 단순 선형회귀모형 $Y_i^* = \gamma_0 + \gamma_1 x_i^* + \eta_i$ 를 적합하였을 때, γ_1 의 최소 제곱추정값과 (x_i^*, y_i^*) 의 표본상관계수를 바르게 연결한 것은? (단, ϵ_i 는 $N(0, \sigma_{\epsilon}^2)$ 을 따르고 서로 독립이며, η_i 는 $N(0, \sigma_{\eta}^2)$ 을 따르고 서로 독립이다)

	γ_1 의 최소제곱추정값	(x_i^*, y_i^*) 의 표본상관계수
①	-0.3	-0.8
②	-0.3	0.8
③	-0.1	-0.8
④	-0.1	0.8

21. $X_1, X_2, \dots, X_{10}$ 이 정규분포 $N(\mu, \sigma^2)$ 으로부터의 임의표본일 때, μ 와 σ^2 에 대한 95 % 신뢰구간에 대한 설명으로 옳은 것만을 모두 고르면?

- ㄱ. σ^2 이 알려져 있는 경우 정규분포를 사용한 μ 에 대한 신뢰 구간의 길이는 $X_1, X_2, \dots, X_{10}$ 의 값에 관계없이 일정하다.
 - ㄴ. σ^2 이 알려져 있지 않은 경우 μ 에 대한 신뢰구간의 길이는 $X_1, X_2, \dots, X_{10}$ 의 값에 관계없이 일정하다.
 - ㄷ. μ 가 알려져 있지 않은 경우 σ^2 에 대한 신뢰구간의 길이는 $X_1, X_2, \dots, X_{10}$ 의 값에 관계없이 일정하다.

- ① ㄱ
- ② ㄱ, ㄴ
- ③ ㄱ, ㄷ
- ④ ㄴ, ㄷ

22. 당뇨병 환자 중 임의로 15명(집단 A)과 10명(집단 B)을 선택하여 집단 A 에게 신약을 처방한 후에 얻은 혈당 개선 효과 자료 $\{x_1, x_2, \dots, x_{15}\}$ 와 집단 B 에게 위약을 처방한 후에 얻은 혈당 개선 효과 자료 $\{y_1, y_2, \dots, y_{10}\}$ 을 다음과 같이 요약하였다.

$$\begin{aligned} \circ \bar{x} &= \frac{1}{15} \sum_{i=1}^{15} x_i \\ \circ \bar{y} &= \frac{1}{10} \sum_{i=1}^{10} y_i \\ \circ s_P &= \sqrt{\frac{1}{23} \left(\sum_{i=1}^{15} (x_i - \bar{x})^2 + \sum_{i=1}^{10} (y_i - \bar{y})^2 \right)} \end{aligned}$$

집단 A 와 B 의 평균 혈당 개선 효과 μ_A 와 μ_B 에 대한 가설 $H_0: \mu_A = \mu_B$ 대 $H_1: \mu_A > \mu_B$ 를 유의수준 α 에서 검정하고자 한다. 이에 대한 설명으로 옳은 것은? (단, 집단 A 와 B 의 혈당 개선 효과는 분산이 동일한 정규분포를 따르고, $t_\alpha(k)$ 는 자유도가 k 인 t 분포의 제 $100 \times (1 - \alpha)$ 백분위수이며, $s_P > 0$ 이다)

- ① $\frac{\sqrt{6}(\bar{x} - \bar{y})}{s_P^2}$ 가 $t_\alpha(23)$ 보다 크면 귀무가설을 기각한다.
- ② $\frac{\sqrt{6}(\bar{x} - \bar{y})}{s_P^2}$ 가 $-t_\alpha(23)$ 보다 작으면 귀무가설을 기각한다.
- ③ $\frac{\sqrt{6}(\bar{x} - \bar{y})}{s_P}$ 가 $t_\alpha(23)$ 보다 크면 귀무가설을 기각한다.
- ④ $\frac{\sqrt{6}(\bar{x} - \bar{y})}{s_P}$ 가 $-t_\alpha(23)$ 보다 작으면 귀무가설을 기각한다.

23. 변수 X 와 Y 의 자료 (x_i, y_i) ($i = 1, 2, \dots, 20$)에서 X 의 표본평균은 1이고 표본분산은 3이다. 이 자료에 단순선형회귀모형 $Y_i = \beta x_i + \epsilon_i$ 를 적합시킨 β 의 최소제곱추정량의 분산은? (단, ϵ_i 는 $N(0, 1)$ 을 따르고 서로 독립이다)

- ① $\frac{1}{57}$
- ② $\frac{1}{60}$
- ③ $\frac{1}{77}$
- ④ $\frac{1}{80}$

24. 확률밀도함수가 다음과 같은 임의표본 X_1, X_2, X_3 의 중앙값을 M 이라 할 때, $P\left(M < \frac{1}{2}\right)$ 의 값은?

$$f(x) = \begin{cases} 2 - 2x, & 0 < x < 1 \\ 0, & \text{그 외} \end{cases}$$

- ① $\frac{5}{32}$
- ② $\frac{7}{16}$
- ③ $\frac{9}{16}$
- ④ $\frac{27}{32}$

25. 반복수가 동일한 이원배치법에서 인자 A 의 수준이 $A1, A2$ 이고 인자 B 의 수준이 $B1, B2$ 인 실험을 통해 얻은 결과의 분산분석표는 다음과 같다.

요인	제곱합	자유도	평균제곱	F -값	p -값
인자 A	784	1	784	26.13	0.0003
인자 B	16	1	16	0.53	0.4792
교호작용($A \times B$)	400	1	400	13.3	0.0033
오차	360	12	30		
전체	1560	15			

인자의 수준 조합에서 반응변수 평균을 나타낸 그림 중 분산분석표의 결과와 부합하는 것은? (단, 유의수준은 5%이다)

