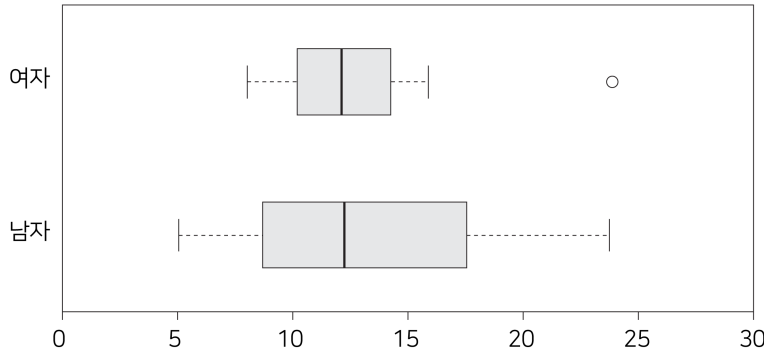


## 통 계 학

1. 다음 그림은 어떤 자료를 성별로 요약한 상자그림을 나타낸다. 이에 대한 설명으로 옳지 않은 것은?



- ① 최솟값은 남자가 여자보다 작다.  
 ② 여자의 자료는 이상점(outlier)을 가진다.  
 ③ 범위(range)는 남자가 여자보다 크다.  
 ④ 사분위수범위(IQR)는 남자가 여자보다 작다.
2. 어느 학생이 지난 학기에 수강한 4과목의 취득점수, 그리고 과목별 전체 평균과 표준편차가 다음과 같다. 각 과목에서 A학점을 받기 위해서는 상위 16% 안에 들어야 한다고 할 때, 이 학생이 수강한 과목 중 A학점을 받은 과목으로만 나열한 것은? (단, 모든 과목의 성적은 각각 정규분포를 따르고,  $Z$ 가  $N(0, 1)$ 을 따를 때  $P(-1 \leq Z \leq 1) = 0.68$ 이다)

수강과목		통계학	미적분학	영어	경제학
취득점수		85	87	90	92
전체	평균	78	82	85	90
	표준편차	5	6	4	3

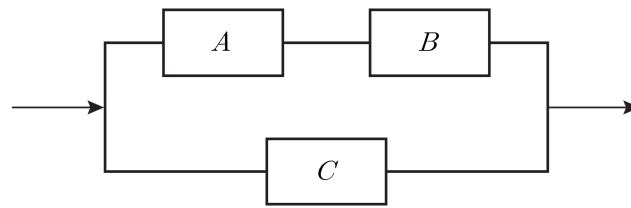
- ① 통계학, 미적분학  
 ② 통계학, 영어  
 ③ 미적분학, 경제학  
 ④ 영어, 경제학
3. 어느 중학교에는 500명의 학생이 재학 중이다. 인터넷 사용 시간에 대한 조사를 수행하는 한 연구자가 성별에 따라 사용 시간에 차이가 큰 것을 알고 전체 학생을 남학생과 여학생으로 구분한 후, 남학생과 여학생의 명부로부터 각각 30명씩 무작위로 추출하였다. 이에 해당하는 확률표본추출방법은?
- ① 집락추출법(cluster sampling)  
 ② 층화추출법(stratified sampling)  
 ③ 할당추출법(quota sampling)  
 ④ 계통추출법(systematic sampling)

4. 분산분석에 대한 다음 설명 중 옳은 것만을 모두 고르면?

- ㄱ. 일원배치법에서 각 수준에서의 반복수가 반드시 같아야 한다.  
 ㄴ. 반복이 있는 이원배치법에서 두 인자  $A, B$  각각의 제곱합 ( $SS_A, SS_B$ )의 자유도는 교호작용 제곱합( $SS_{A \times B}$ )의 자유도보다 항상 작다.  
 ㄷ. 이원배치법에서 교호작용의 존재 여부를 파악하고자 할 경우에는 반복이 있는 이원배치법을 사용해야 한다.

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

5. 다음 그림은 세 개의 스위치  $A, B, C$ 로 구성된 전기전달시스템을 보여 준다. 스위치  $A$ 가 ON일 확률은 0.8, 스위치  $B$ 가 ON일 확률은 0.7, 스위치  $C$ 가 ON일 확률은 0.6이라 하자. 각각의 스위치가 독립적으로 작동한다고 할 때, 이 시스템에서 전기가 전달될 확률은?



- ① 0.784  
 ② 0.804  
 ③ 0.824  
 ④ 0.844

6. 연속형 확률변수  $X$ 와  $Y$ 의 결합확률밀도함수가 다음과 같을 때, 옳은 것은? (단,  $a$ 는 상수이다)

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

- ①  $P(X \leq \frac{1}{2}) = \frac{3}{4}$   
 ②  $P(Y \leq \frac{1}{3}) = \frac{1}{9}$   
 ③  $P(X > \frac{1}{4}) = \frac{1}{16}$   
 ④  $P(Y \leq \frac{1}{5}) = \frac{24}{25}$

7. 두 확률변수  $X, Y$ 의 결합확률분포가 다음과 같다.

$X \backslash Y$	0	1
1	$\frac{1}{4}$	$\frac{1}{4}$
2	$\frac{1}{2}$	0

확률변수  $W$ 와  $Z$ 를  $W = -4X + 3, Z = 8Y - 3$ 이라 할 때, 공분산  $\text{Cov}(W, Z)$ 의 값은?

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

8. 주사위를 1회 던질 때 2의 배수가 나오는 사건을  $A$ , 3의 배수가 나오는 사건을  $B$ , 2의 배수가 나오지 않는 사건을  $C$ , 3의 배수가 나오지 않는 사건을  $D$ 라고 하자. 다음 중 두 사건이 서로 독립인 경우만을 모두 고르면?

- ㄱ.  $A, B$   
ㄴ.  $B, C$   
ㄷ.  $C, D$

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

9. 평균이  $\mu$ 이고 분산이  $\sigma^2$ 인 정규모집단으로부터 추출된 크기 2인 확률표본을  $X_1, X_2$ 라고 하자.  $X_1$ 과  $X_2$ 의 표준화된 확률변수를 각각  $Z_1$ 과  $Z_2$ 라고 할 때, 다음 중 옳지 않은 것은?

- ①  $E(Z_1) = E(Z_2)$
- ②  $Var(X_1) = 2 \times Var\left(\frac{X_1 + X_2}{2}\right)$
- ③  $Var(Z_1^2) = 1$
- ④  $E(Z_1^2 + Z_2^2) = 2$

10. 표준편차가 6인 정규모집단에서 모평균( $\mu$ )에 대한 가설  $H_0 : \mu = 10$  대  $H_1 : \mu > 10$ 을 검정하고자 한다. 크기  $n$ 의 확률표본으로부터 구해진 표본평균을  $\bar{X}$ 라고 하고, 이 검정의 기각역으로  $R : \bar{X} \geq 11$ 을 사용하고자 한다.  $\mu = 13$ 일 때, 검정력이 0.95 이상이 되도록 하는 최소 표본의 크기  $n$ 을 구하는 부등식은? (단,  $z_\alpha$ 는 표준정규분포의 제  $100 \times (1 - \alpha)$  백분위수를 나타내고,  $z_{0.05} = 1.645, z_{0.025} = 1.96$ 이다)

- ①  $n \geq (3 \times 1.645)^2$
- ②  $n \geq (3 \times 1.96)^2$
- ③  $n \geq (6 \times 1.645)^2$
- ④  $n \geq (6 \times 1.96)^2$

11. 평균이  $\mu$ 이고 분산이  $\sigma^2$ 인 모집단으로부터 추출된 크기 5인 확률 표본을  $X_1, X_2, \dots, X_5$ 라고 하자. 다음에 제시된  $\mu$ 에 대한 추정량 중에서 불편추정량인면서 동시에 분산이 제일 작은 추정량은? (단,  $\sigma^2 > 0$ 이다)

$$T_1 = \frac{X_1 + X_5}{2}, \quad T_2 = \frac{2X_1 + 2X_2 - X_3}{3},$$
$$T_3 = \frac{X_1 + X_2 + X_5}{4}, \quad T_4 = X_3 + X_4 - X_5$$

- ①  $T_1$
- ②  $T_2$
- ③  $T_3$
- ④  $T_4$

12.  $X_1, X_2, \dots, X_5$ 를  $(\theta, \theta + 1)$ 에서 균일분포(uniform distribution)를 따르는 모집단으로부터 추출된 크기 5인 확률표본이라 하고, 표본평균을  $\bar{X}$ 라고 하자. 모수  $\theta$ 에 대한 추정량으로  $\hat{\theta} = \bar{X} - \frac{1}{2}$ 을 사용할 때, 이 추정량의 편의(bias)와 평균제곱오차(MSE)를 바르게 연결한 것은?

	편의	평균제곱오차
①	0	$\frac{1}{60}$
②	0	$\frac{1}{30}$
③	$\frac{1}{2}$	$\frac{1}{60}$
④	$\frac{1}{2}$	$\frac{1}{30}$

13. 관측 자료  $(x_i, y_i) (i=1,2,\cdots,50)$ 를 이용하여 단순선형회귀모형  $Y_i = \beta_0 + \beta_1 x_i + \epsilon_i$ 를 최소제곱법으로 추정한 결과  $\hat{\beta}_0 = 200$ ,  $\hat{\beta}_1 = -0.4$ 를 얻었다. 설명변수의 평균이 150, 표준편차가 80이고, 추정된 회귀직선의 결정계수( $R^2$ )가 0.81일 때, 이에 대한 설명으로 옳은 것만을 모두 고르면? (단,  $\epsilon_i$ 는  $N(0, \sigma^2)$ 을 따르고 서로 독립이다)

- ㄱ. 설명변수와 반응변수 간의 표본상관계수는  $-0.9$ 이다.
- ㄴ. 반응변수의 평균은 140이다.
- ㄷ.  $x=100$ 일 때, 반응변수의 추정값은 160이다.

- ① ㄱ
- ② ㄱ, ㄴ
- ③ ㄴ, ㄷ
- ④ ㄱ, ㄴ, ㄷ

14. 모수가  $\lambda$ 인 포아송 분포를 따르는 확률변수( $X$ )의 확률질량함수는 다음과 같다.

$$f(x) = \begin{cases} \frac{e^{-\lambda} \lambda^x}{x!}, & x=0,1,2,\cdots \\ 0, & \text{그 외} \end{cases} \quad (\lambda > 0)$$

$X_1, X_2$ 를 서로 독립이며 각각 모수가  $\lambda_1 = 3, \lambda_2 = 2$ 인 포아송 분포를 따르는 확률변수라고 할 때,  $P(X_1 + 2X_2 = 2)$ 의 값은?

- ①  $\frac{9}{2}e^{-5}$
- ②  $\frac{11}{2}e^{-5}$
- ③  $\frac{13}{2}e^{-5}$
- ④  $\frac{25}{2}e^{-5}$

15. 다음은 어떤 자료에 대해 단순선형회귀모형  $Y_i = \beta_0 + \beta_1 x_i + \epsilon_i$  ( $i=1,2,\cdots,16$ )를 적합한 결과와 그에 따른 해석이다. (가), (나)에 들어갈 내용으로 바르게 연결한 것은? (단,  $\epsilon_i$ 는  $N(0, \sigma^2)$ 을 따르고 서로 독립이다. 또한  $t_\alpha(k)$ 는 자유도가  $k$ 인  $t$ 분포의 제 $100 \times (1-\alpha)$  백분위수를 나타내고,  $t_{0.025}(14) = 2.15$ ,  $t_{0.05}(14) = 1.76$ 이다)

계수	추정회귀계수	표준오차	$t$ -값	$p$ -값
$\beta_0$	120	40	3	0.0096
$\beta_1$	1.8	0.5	3.6	0.0029

$\beta_0$ 에 대한 95% 신뢰구간은 (가)이고, 가설  $H_0 : \beta_1 \leq 1$  대  $H_1 : \beta_1 > 1$ 에 대한 유의수준 5%의 검정 결과 귀무가설을 기각할 수 (나).

- |  | (가) | (나) |
|--|-----|-----|
| ① $(120 - 1.76 \times 40, 120 + 1.76 \times 40)$ |     | 있다  |
| ② $(120 - 1.76 \times 40, 120 + 1.76 \times 40)$ |     | 없다  |
| ③ $(120 - 2.15 \times 40, 120 + 2.15 \times 40)$ |     | 있다  |
| ④ $(120 - 2.15 \times 40, 120 + 2.15 \times 40)$ |     | 없다  |

16. 남자 110명, 여자 110명을 각각 임의추출하여 가장 지지하는 정당을 조사한 결과가 다음과 같다. 이 조사에서 무응답이나 중복 응답은 없다고 한다.

성별 \ 정당	A	B	C	계
남자	55	30	25	110
여자	45	50	15	110

정당별 지지율이 성별에 따라 다른지를 알아보기 위해 카이제곱 검정을 실시하였다. 이 검정에서  $p$ -값이 속하는 범위는? (단,  $\chi^2_\alpha(k)$ 는 자유도가  $k$ 인 카이제곱분포의 제 $100 \times (1-\alpha)$  백분위수를 나타내고,  $\chi^2_{0.01}(2) = 9.21$ ,  $\chi^2_{0.05}(2) = 5.99$ ,  $\chi^2_{0.1}(2) = 4.61$ ,  $\chi^2_{0.01}(5) = 15.09$ ,  $\chi^2_{0.05}(5) = 11.07$ ,  $\chi^2_{0.1}(5) = 9.24$ 이다)

- ①  $p\text{-값} < 0.01$
- ②  $0.01 \leq p\text{-값} < 0.05$
- ③  $0.05 \leq p\text{-값} < 0.1$
- ④  $p\text{-값} \geq 0.1$

17. 어느 제약회사에서 코로나 치료제를 개발하기 위해 40명의 코로나 환자를 대상으로 4가지 약품에 대해 임상 시험을 진행하였다. 이 시험에서 얻은 자료로 분산분석표를 작성한 결과가 다음과 같을 때, (가) ~ (다)의 값으로 바르게 연결한 것은?

요인	제곱합	자유도	평균제곱	F-값
처리	30	(가)		(다)
오차		36	(나)	
계	210			

	(가)	(나)	(다)
①	3	5	2
②	3	6.7	1.5
③	4	5	2
④	4	6.7	1.5

18. 두 모집단의 평균을 비교하기 위해 가설  $H_0: \mu_1 = \mu_2$  대  $H_1: \mu_1 \neq \mu_2$ 에 대한 검정을 실시하고자 한다. 두 모집단으로부터 각각 임의추출된 10, 12개의 자료에 대해 두 가지 검정법을 적용하였다. 방법 A로 독립인 두 모집단에 대한 t-검정, 방법 B로 수준이 2인 일원배치법을 적용할 때, 이에 대한 설명으로 옳은 것만을 모두 고르면? (단, 두 모집단은 서로 독립이며 분산이 동일한 정규분포를 따른다)

- ㄱ. 방법 A에서의 검정통계량 t-값의 제곱은 방법 B에서의 검정통계량 F-값과 같다.

ㄴ. 두 검정 방법 A와 B에서의 p-값은 같다.

ㄷ. 방법 A에서의 검정통계량 t의 자유도는 방법 B에서의 검정통계량 F의 분자의 자유도와 같다.

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

19. 어떤 화력발전소에서 배출되는 미세먼지의 일별 배출량은 표준편차가 1인 정규분포를 따르는 것으로 알려져 있다. 임의로 추출한 36일간의 미세먼지 일별 배출량의 평균과 실제 모평균의 차이가 0.1 이상일 확률을 나타내는 식은? (단, Z는  $N(0, 1)$ 을 따르고, 미세먼지의 일별 배출량은 서로 독립이다)

- ①  $P(|Z| \leq 0.6)$   
②  $2 \times P(Z \geq 0.6)$   
③  $P(|Z| \leq 0.1)$   
④  $2 \times P(Z \geq 0.1)$

20. 변수 X와 Y 사이의 관계를 단순선형회귀모형  $Y_i = \beta_0 + \beta_1 x_i + \epsilon_i$  ( $i = 1, 2, \dots, n$ )로 분석하고자 한다. 이를 위해  $\beta_0, \beta_1$ 의 최소제곱추정량  $\hat{\beta}_0, \hat{\beta}_1$ 을 각각 구하고 잔차  $e_i = Y_i - \hat{\beta}_0 - \hat{\beta}_1 x_i$ 를 구하였다. 이와 관련된 설명으로 옳지 않은 것은? (단,  $\epsilon_i$ 는  $N(0, \sigma^2)$ 을 따르고 서로 독립이다)

- ①  $\hat{\beta}_0, \hat{\beta}_1$ 은 각각  $\beta_0, \beta_1$ 의 불편추정량이다.  
② 가설  $H: \beta_1 = 0$ 은 X와 Y 간에 어떠한 관계성도 없음을 의미한다.  
③ 잔차들의  $x_i$ 에 대한 가중합  $\sum_{i=1}^n x_i e_i$ 는 0이다.  
④  $\hat{\beta}_1$ 의 분산은  $\frac{\sigma^2}{\sum_{i=1}^n (x_i - \bar{x})^2}$ 이다.

21. 0과 1로만 이루어져 있고 1의 비율이 p인 모집단으로부터 5개의 표본을 임의추출한 결과, 4개가 1로 관측되었다. 가설  $H_0: p = \frac{2}{3}$  대  $H_1: p > \frac{2}{3}$ 에 대한 검정에서 유의확률은?

- ①  $\frac{1}{3^5}$   
②  $\frac{11}{3^5}$   
③  $\frac{80}{3^5}$   
④  $\frac{112}{3^5}$

22. 다음은 단순선형회귀모형  $Y_i = \beta_0 + \beta_1 x_i + \epsilon_i$  ( $i = 1, 2, \dots, n$ )를 적합하기 위해 어떤 자료를 요약한 결과이다. 이에 대한 설명으로 옳지 않은 것은? (단,  $\epsilon_i$ 는  $N(0, \sigma^2)$ 을 따르고 서로 독립이다)

$$n = 12, \bar{x} = 1.5, \bar{y} = 8,$$
$$\sum_{i=1}^{12} (x_i - \bar{x})^2 = 10, \sum_{i=1}^{12} (y_i - \bar{y})^2 = 90, \sum_{i=1}^{12} (x_i - \bar{x})(y_i - \bar{y}) = 20$$

- ①  $\beta_1$ 의 최소제곱추정값은 2이다.  
②  $\beta_0$ 의 최소제곱추정값은 5이다.  
③ 잔차의 제곱합은 40이다.  
④ 오차분산의 추정값은 5이다.

23. 어느 도시에서 특정 지역축제의 개최에 대한 찬성률( $p$ )을 알아보기 위해 시민 100명을 대상으로 작년에 처음 실시한 여론조사에서, 찬성률의 추정에 95% 신뢰수준을 사용하였다. 올해 실시 예정인 동일한 여론조사에서, 오차한계를 포함한 다른 조건은 모두 작년과 동일하게 하되 표본 수를 줄이기 위해 신뢰수준만 90%로 바꾸고자 한다. 올해 조사해야 하는 최소 표본의 수는? (단,  $z_{\alpha}$ 는 표준정규분포의 제 $100 \times (1 - \alpha)$  백분위수를 나타내고, 계산의 편의상  $z_{0.1} = 1.3$ ,  $z_{0.05} = 1.6$ ,  $z_{0.025} = 2$ 를 사용하기로 한다)

① 43명  
② 64명  
③ 87명  
④ 92명

24. 다음은 어느 대학의 A학과와 B학과 간에 통계학 성적에 차이가 있는지를 알아보기 위해 두 학과에서 각각 11명, 16명을 임의추출하여 조사한 결과이다. 두 학과의 통계학 성적 간에 차이가 있는지를 검정하기 위한  $t$ -검정통계량( $t$ )의 값과 유의수준 5%에서의 기각역으로 바르게 연결한 것은? (단, 각 학과의 통계학 성적은 서로 독립이며 정규분포를 따르고 분산은 같다. 또한,  $t_{\alpha}(k)$ 는 자유도가  $k$ 인  $t$ 분포의 제 $100 \times (1 - \alpha)$  백분위수이다)

학과	표본 수	평균	표준편차
A	11	70	6
B	16	65	4

- $t$ -검정통계량( $t$ )의 값

기각역
- ①  $\frac{70 - 65}{\sqrt{24 \times \left( \frac{1}{11} + \frac{1}{16} \right)}}$

$|t| \geq t_{0.025}(26)$
- ②  $\frac{70 - 65}{\sqrt{\frac{36}{11} + \frac{16}{16}}}$

$|t| \geq t_{0.025}(26)$
- ③  $\frac{70 - 65}{\sqrt{24 \times \left( \frac{1}{11} + \frac{1}{16} \right)}}$

$|t| \geq t_{0.025}(25)$
- ④  $\frac{70 - 65}{\sqrt{\frac{36}{11} + \frac{16}{16}}}$

$|t| \geq t_{0.025}(25)$

25. 새로 제작된 3개의 동전을 동시에 던지는 실험을 독립적으로 80회 실시하였다. 다음 표는 ‘각 실험에서 앞면이 나온 수( $X$ )’의 빈도(도수)를 나타낸다.

$X$	$H_0$ 하의 기대도수	관측도수
0	$E_1$	8
1	$E_2$	27
2	$E_3$	34
3	$E_4$	11
계	80	80

- 가설 ‘ $H_0$ :  $X$ 가 이항분포  $B(3, \frac{1}{2})$ 을 따른다’에 대한 카이제곱검정을 실시할 때, 다음 설명 중 옳지 않은 것은? (단,  $\chi^2_{\alpha}(k)$ 는 자유도가  $k$ 인 카이제곱분포의 제 $100 \times (1 - \alpha)$  백분위수이다)
- ①  $E_2$ 는 30이다.  
②  $H_0$ 하에서  $P(X=2)$ 는  $\frac{3}{8}$ 이다.  
③ 검정통계량의 값은  $\frac{5}{3}$ 이다.  
④ 유의수준 5%인 검정의 임계값은  $\chi^2_{0.05}(3)$ 이다.