

## 통 계 학

문 1. 다음 글이 설명하고 있는 표본추출방법은?

어떤 지역에 새로운 상점을 열기 전에 잠재적인 구매 수요를 조사하기 위하여 그 지역에 거주하는 주민들을 임의로 추출하고자 한다. 조사를 위한 이동 비용을 최소화하기 위하여 먼저 이 지역을 여러 개의 블록으로 나눈다. 그리고 그 중에서 몇 개의 블록을 임의로 추출하여 그 블록에 거주하는 모든 주민들의 구매 수요를 조사한다. 블록을 구성할 때 가급적이면 같은 블록 내에서는 구매 수요의 이질성을 높이고 다른 블록들 간에서는 구매 수요의 동질성을 높이도록 하였다.

- ① 단순임의추출법(simple random sampling)
- ② 층화추출법(stratified sampling)
- ③ 계통추출법(systematic sampling)
- ④ 군집(집락)추출법(cluster sampling)

문 2. 두 확률변수  $X$ 와  $Y$ 에 대하여  $Var(X+Y)=6$ 이고  $Var(X-Y)=10$ 일 때,  $X$ 와  $Y$ 의 공분산은?

- ① -2
- ② -1
- ③ 1
- ④ 2

문 3. 두 확률변수  $X$ 와  $Y$ 에 대한 관측 쌍  $(x_1, y_1), (x_2, y_2), \dots, (x_{10}, y_{10})$ 으로부터 계산된 통계량이 다음과 같을 때, 표본상관계수(sample correlation coefficient)는?

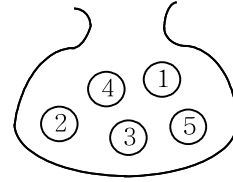
$$\sum_{i=1}^{10} (x_i - \bar{x})^2 = 4, \sum_{i=1}^{10} (y_i - \bar{y})^2 = 25, \sum_{i=1}^{10} (x_i - \bar{x})(y_i - \bar{y}) = 9$$

- ① 0.03
- ② 0.30
- ③ 0.81
- ④ 0.90

문 4. 어느 지역의 초등학교 키는 정규분포를 따르고, 이 지역 초등학교 100명을 임의 추출하여 조사한 키의 표본평균과 표본표준편차는 각각 140cm와 15cm였다. 이 지역 초등학교 키의 모평균에 대한 95% 신뢰구간은? (단,  $z_{\alpha}$ 는 표준정규분포의  $(1-\alpha) \times 100$ 번째 백분위수를 나타낸다)

- ①  $140 \pm z_{0.025} \times \frac{15}{\sqrt{100}}$
- ②  $140 \pm z_{0.05} \times \frac{15}{\sqrt{100}}$
- ③  $140 \pm z_{0.025} \times 15$
- ④  $140 \pm z_{0.05} \times 15$

문 5. 다음 그림과 같이 어떤 바구니에 1, 2, 3, 4, 5의 숫자가 각각 하나씩 적혀 있는 5개의 공이 들어있다. 이 바구니에서 세 개의 공을 임의로 비복원추출할 때, 이 세 개의 공에 적혀 있는 숫자의 중앙값(median)이 3일 확률은?



- ① 0.2
- ② 0.4
- ③ 0.6
- ④ 0.8

문 6. 어떤 사람이 20분 당 평균 3건의 문자메시지를 받는다. 이 사람이 1시간 동안 받는 문자메시지의 수가 포아송분포를 따르는 확률변수  $X$ 일 때,  $X$ 의 변동계수(coefficient of variation)는? (단, 변동계수의 단위는 백분율(%)이다)

- ①  $\frac{1}{3} \times 100$
- ②  $\frac{1}{\sqrt{3}} \times 100$
- ③  $\sqrt{3} \times 100$
- ④  $3 \times 100$

문 7. 행운권 100장을 임의로 추첨하여 1등 1장은 100,000원, 2등 2장은 30,000원, 3등 3장은 10,000원의 상금을 받는다. 이 행운권을 사기 위해서는 한 장에 1,000원을 내야한다. 어떤 사람이 이 행운권을 한 장 샀을 때의 기대이익금액은? (단, 1, 2, 3등 이외의 등수는 없다)

- ① 900원
- ② 1,000원
- ③ 1,400원
- ④ 1,900원

문 8.  $X_1, X_2, \dots, X_n$ 은 평균이  $\mu$ 이고 분산이 1인 정규분포로부터의 임의표본(random sample)이다. 표본의 크기  $n$ 이 증가할 때, 옳은 것을 모두 고르면? (단, 표본의 크기가 증가해도 다른 조건은 변하지 않는다)

ㄱ.  $Var(X_n)$ 은 작아진다.

ㄴ. 모평균  $\mu$ 에 대한  $(1-\alpha) \times 100$  % 신뢰구간의 길이는 짧아진다.

ㄷ. 주어진 모평균  $\mu_0$ 에 대한 가설  $H_0: \mu = \mu_0$  대  $H_1: \mu < \mu_0$ 의 검정력(power of test)은 증가한다.

- ① ㄴ
- ② ㄱ, ㄷ
- ③ ㄴ, ㄷ
- ④ ㄱ, ㄴ, ㄷ

- 문 9. 인자 A는 세 가지 수준을 갖고 인자 B는 네 가지 수준을 갖는, 반복이 없는 이원배치 분산분석법을 적용하여 얻은 분산분석표의 일부가 다음과 같다. ㉠의 값은?

| 요인   | 제곱합 | F 비 |
|------|-----|-----|
| 인자 A | 4   | ㉠   |
| 인자 B | 9   |     |
| 오차   |     |     |
| 합계   | 16  |     |

- ① 0.44  
② 0.67  
③ 4.00  
④ 6.00
- 문 10. 원점을 지나는 단순선형회귀모형  $Y_i = \beta X_i + \epsilon_i$ 에서 최소제곱법(method of least squares)에 의해 추정된 회귀직선  $\hat{Y}_i = \hat{\beta} X_i$ 에 대한 설명으로 옳지 않은 것은? (단,  $i = 1, 2, \dots, n$ 이고, 잔차는  $e_i = Y_i - \hat{Y}_i$  이다)

- ① 잔차들의 합인  $\sum_{i=1}^n e_i$ 는 0이 아닐 수도 있다.

- ②  $\beta$ 에 대한 최소제곱추정량  $\hat{\beta}$  은  $\frac{\sum_{i=1}^n X_i Y_i}{\sum_{i=1}^n X_i^2}$  이다.

- ③ 잔차제곱합인  $\sum_{i=1}^n e_i^2$ 은  $(n-1)$ 의 자유도를 가진다.

- ④ 추정된 회귀직선은 항상  $X_i$ 들과  $Y_i$ 들의 표본평균점  $(\bar{X}, \bar{Y})$ 를 지난다.

- 문 11. 다음은 어느 대학교의 통계학개론 강좌를 수강한 학생 30명의 기말고사 점수에 대한 줄기-잎 그림(stem-and-leaf plot)이다. 이 자료의 중앙값(median)과 범위(range)는?

|   |          |
|---|----------|
| 2 | 2        |
| 3 | 26       |
| 4 | 18       |
| 5 | 258      |
| 6 | 2259     |
| 7 | 12334555 |
| 8 | 224599   |
| 9 | 1145     |

- |   | 중앙값 | 범위 |
|---|-----|----|
| ① | 73  | 69 |
| ② | 73  | 73 |
| ③ | 75  | 69 |
| ④ | 75  | 73 |

- 문 12. 확률변수  $X_1$ 은 평균이 30이고 표준편차가 5이며, 확률변수  $X_2$ 는 평균이 20이고 표준편차가 3이다.  $X_1$ 과  $X_2$ 가 서로 독립일 때, 확률변수  $Y = 2X_1 - X_2 + 4$ 의 평균과 분산은?

|   | 평균 | 분산  |
|---|----|-----|
| ① | 40 | 109 |
| ② | 40 | 113 |
| ③ | 44 | 109 |
| ④ | 44 | 113 |

- 문 13. 두 확률변수  $X$ 와  $Y$ 가 서로 독립이고 다음을 만족할 때, 옳지 않은 것은?

$$E(X) = 3, E(Y) = 2, Var(X) = 4, Var(Y) = 2$$

- ①  $E(XY) = 6$   
②  $E(X^2) > E(Y^2)$   
③  $Var(X - Y) = 2$   
④  $Cov(2X + 3, 3Y + 2) = 0$

- 문 14. 다음은 어느 학과 학생들의 가족 구성원 수를 조사하여 만든 도수분포표이다. ㉠ + ㉡ + ㉢ + ㉣의 값은?

| 가족 구성원 수 | 도수(명) | 상대도수 |
|----------|-------|------|
| 2        | ㉠     | 0.2  |
| 3        | 10    | ㉢    |
| 4        | ㉡     | 0.2  |
| 5        | 2     | ㉣    |

- ① 6.6  
② 8.6  
③ 10.6  
④ 12.6

- 문 15. 사건 A가 일어날 확률은 0.5이고 사건 B가 일어날 확률은 0.2이다. 사건 A와 사건 B가 서로 독립일 때, 사건  $A^C \cup B^C$ 가 일어날 확률은?

- ① 0.6  
② 0.7  
③ 0.8  
④ 0.9

- 문 16. 연속형 확률변수  $X$ 의 확률밀도함수  $f(x)$ 가 다음과 같다.  $P(X \leq t) = 0.5$ 를 만족하는  $t$ 의 값은?

$$f(x) = \begin{cases} \frac{x}{2}, & 0 \leq x \leq 2 \\ 0, & x < 0 \text{ 또는 } x > 2 \end{cases}$$

- ① 0  
② 1  
③  $\sqrt{2}$   
④ 2

문 17. 어떤 동전을 100번 던졌더니 앞면이 60번 나왔다. 이 결과를 이용하여 ‘이 동전의 앞면이 나올 확률은 0.5이다’라는 귀무가설을 검정하기 위하여 피어슨 카이제곱 검정을 실시할 때, 검정통계량 값은? (단, 이 동전은 앞면 또는 뒷면만 나온다)

- ① 4  
② 6  
③ 40  
④ 60

문 18. 표본의 크기가 10인 어느 자료로부터 단순선형회귀모형  $Y_i = \alpha + \beta X_i + \epsilon_i$ ,  $i = 1, 2, \dots, 10$ 을 적용하여 얻은 분산분석표의 일부가 다음과 같다. (단,  $\epsilon_i$ 는 서로 독립이며 평균이 0이고 분산이  $\sigma^2$ 인 정규분포를 따른다)

| 요인 | 제곱합 | 자유도 | 평균제곱 | p-값    |
|----|-----|-----|------|--------|
| 회귀 |     |     | 10   | 0.0021 |
| 오차 | 4   |     | Ⓐ    |        |
| 합계 |     | 9   |      |        |

이에 대한 설명으로 옳은 것을 모두 고르면?

- ㄱ. Ⓐ의 값은 0.5이다.  
 ㄴ. 이 회귀모형의 결정계수(coefficient of determination)는  $\frac{10}{15}$ 이다.  
 ㄷ. 유의수준 1%에서 ‘이 회귀모형이 유의하지 않다’는 귀무가설을 기각한다.

- ① ㄷ  
② ㄱ, ㄴ  
③ ㄱ, ㄷ  
④ ㄱ, ㄴ, ㄷ

문 19. 단순선형회귀모형  $Y_i = \alpha + \beta X_i + \epsilon_i$ 에서  $\epsilon_i$ 는 서로 독립이며 평균이 0이고 분산이  $\sigma^2$ 인 정규분포를 따를 때, 최소제곱법(method of least squares)에 의해 추정된 회귀직선  $\hat{Y}_i = \hat{\alpha} + \hat{\beta}X_i$ 에 대한 설명으로 옳은 것을 모두 고르면? (단,  $i = 1, 2, \dots, n$ 이고, 표본의 크기  $n$ 은 4 이상이다)

- ㄱ. 최소제곱추정량  $\hat{\alpha}$ 와  $\hat{\beta}$ 는 각각  $\alpha$ 와  $\beta$ 의 불편추정량이다.  
 ㄴ. 잔차를  $e_i = Y_i - \hat{Y}_i$ 이라 하면, 잔차들의  $X_i$ 에 대한 가중합은 0이다. 즉,  $\sum_{i=1}^n X_i e_i = 0$ 이다.  
 ㄷ.  $\sigma^2$ 의 불편추정량은  $\sum_{i=1}^n (Y_i - \hat{Y}_i)^2$ 이다.

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

문 20. 평균이  $\mu$ 이고 분산이 4인 정규분포로부터  $n$ 개의 표본을 임의 추출하여  $\mu$ 에 대한 가설  $H_0: \mu = 100$  대  $H_1: \mu > 100$ 을 검정하고자 한다. 귀무가설  $H_0$ 에 대한 기각역이 ‘ $\bar{X} > 104$ ’일 경우, 실제 모평균  $\mu$ 가 108일 때의 검정력(power of test)을 나타내는 식은? (단,  $\bar{X}$ 는 표본평균이고,  $Z$ 는 표준정규분포를 따르는 확률변수이다)

- ①  $P(Z \leq \frac{104-100}{2/\sqrt{n}})$   
 ②  $P(Z > \frac{104-100}{2/\sqrt{n}})$   
 ③  $P(Z \leq \frac{104-108}{2/\sqrt{n}})$   
 ④  $P(Z > \frac{104-108}{2/\sqrt{n}})$