

# 통 계 학

문 1. 다음 사례에 해당하는 표본추출방법은?

2,000 가구에 1부터 2,000까지 서로 다른 일련번호를 하나씩 부여한 후, 일련번호가 1부터 40인 가구 중에서 무작위로 한 가구를 추출하고, 그 일련번호로부터 매 40 번째 일련번호에 해당하는 가구를 추출하여 총 50 가구를 표본으로 선정하였다.

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

문 2. 단순선형회귀모형  $Y_i = \beta_0 + \beta_1 X_i + \epsilon_i$ 에서 추정된 회귀직선이  $\hat{Y}_i = 2 - 0.5X_i$ 이고 결정계수(coefficient of determination)가 0.81 일 때, 두 변수  $X, Y$ 의 표본상관계수는? (단,  $i = 1, 2, \dots, n$  이고,  $\epsilon_i$ 는 서로 독립이며 평균이 0이고 분산이  $\sigma^2$ 인 정규분포를 따른다)

- ① - 0.9
- ② - 0.5
- ③ 0.5
- ④ 0.9

문 3. 표본공간  $S$ 에서 정의된 두 사건  $A, B$ 에 대하여  $P(A) = 0.5$ ,  $P(A \cup B) = 0.7$ 일 때, 옳지 않은 것은? (단,  $A^c$ 은  $A$ 의 여사건을 나타낸다)

- ①  $A$ 와  $B$ 가 서로 독립이면  $P(B) = 0.4$ 이다.
- ②  $A$ 와  $B$ 가 서로 독립이면  $P(A^c \cap B) = 0.2$ 이다.
- ③  $A$ 와  $B$ 가 서로 배반이면  $P(B) = 0.2$ 이다.
- ④  $P(B|A) = 0.5$ 이면  $P(B) = 0.25$ 이다.

문 4. 확률변수  $X$ 는 평균이 50이고 분산이  $\sigma^2$ 인 정규분포를 따르고, 확률변수  $Y$ 는 평균이 60이고 분산이  $9\sigma^2$ 인 정규분포를 따른다고 할 때,  $P(X > 60) = P(Y < c)$ 를 만족하는  $c$ 의 값은? (단,  $\sigma^2 > 0$ )

- ① 10
- ② 20
- ③ 30
- ④ 40

문 5. 다음은 직업을 A, B, C로 구분할 때 아버지의 직업에서 아들의 직업으로 전이될 조건부 확률을 나타낸 것이다. 아버지의 직업이 A, B, C일 확률은 각각 0.1, 0.4, 0.5라고 하자. 아들의 직업이 A일 때, 그의 아버지의 직업이 A일 확률은?

| 아버지 | 아들   |      |      |
|-----|------|------|------|
|     | A    | B    | C    |
| A   | 0.45 | 0.45 | 0.1  |
| B   | 0.1  | 0.7  | 0.2  |
| C   | 0.05 | 0.5  | 0.45 |

- ①  $\frac{7}{22}$
- ②  $\frac{9}{22}$
- ③  $\frac{1}{2}$
- ④  $\frac{13}{22}$

문 6. 두 전기회사 A, B에서 생산하는 전구의 수명시간은 모분산이 각각  $\sigma_A^2, \sigma_B^2$ 인 정규분포를 따른다고 한다. 두 회사 A, B의 전구를 각각 10개, 15개 무작위로 추출하여 수명시간의 표본분산  $s_A^2, s_B^2$ 을 구하였다. 표본분산의 비를  $f = \frac{s_A^2}{s_B^2}$ 이라고 할 때, 모분산의

비  $\frac{\sigma_A^2}{\sigma_B^2}$ 에 대한  $(1 - \alpha) \times 100\%$  신뢰구간은? (단,  $F_\alpha(m, n)$ 은 분자의 자유도가  $m$ , 분모의 자유도가  $n$ 인  $F$ 분포의 제  $(1 - \alpha) \times 100$  백분위수를 나타낸다)

- ①  $(f \cdot \frac{1}{F_{\alpha/2}(9, 14)}, f \cdot F_{\alpha/2}(14, 9))$
- ②  $(f \cdot F_{1-\alpha/2}(9, 14), f \cdot F_{\alpha/2}(9, 14))$
- ③  $(f \cdot \frac{1}{F_{\alpha/2}(14, 9)}, f \cdot F_{\alpha/2}(14, 9))$
- ④  $(f \cdot F_{1-\alpha/2}(14, 9), f \cdot F_{\alpha/2}(9, 14))$

문 7. 다음은 각각 세 가지 수준을 갖는 인자 A와 B의 아홉 가지 수준 조합마다 3회 반복 측정하여 얻은 분산분석표의 일부일 때, 옳지 않은 것은? (단,  $F_\alpha(m, n)$ 은 분자의 자유도가  $m$ , 분모의 자유도가  $n$ 인  $F$ 분포의 제  $(1 - \alpha) \times 100$  백분위수를 나타내고,  $F_{0.05}(2, 18) = 3.55, F_{0.05}(3, 18) = 3.16, F_{0.05}(4, 18) = 2.93$ 이다)

| 요인   | 제곱합 | 자유도 | 평균제곱 | F-값 |
|------|-----|-----|------|-----|
| 인자 A | 144 |     | 72   |     |
| 인자 B | 44  |     |      | 1.1 |
| 교호작용 | 52  |     | 13   |     |
| 잔차   | 360 |     |      |     |
| 계    | 600 |     |      |     |

- ① 인자 A의 주효과가 유의수준 5%에서 유의하다.
- ② 인자 B의 주효과가 유의수준 5%에서 유의하지 않다.
- ③ 인자 A와 인자 B의 교호작용효과는 유의수준 5%에서 유의하다.
- ④ 잔차평균제곱의 값은 20이다.

문 8. 표본의 크기가  $n$ 인 자료  $(X_i, Y_i)$ 에 단순선형회귀모형을 적용하여 추정된 회귀직선이  $\hat{Y}_i = 2 + 4X_i$ 이다.  $Z_i = \frac{1}{2}Y_i + 1$ 이라고 할 때 자료  $(X_i, Z_i)$ 에 단순선형회귀모형을 적용하여 추정된 회귀직선이  $\hat{Z}_i = \hat{\beta}_0 + \hat{\beta}_1 X_i$ 라면, 옳지 않은 것은? (단,  $i = 1, 2, \dots, n$ 이고,  $\bar{Y} = \frac{1}{n} \sum_{i=1}^n Y_i, \bar{Z} = \frac{1}{n} \sum_{i=1}^n Z_i$ 이다)

- ①  $\hat{\beta}_0 = 2$
- ②  $\hat{\beta}_1 = 2$
- ③  $\sum_{i=1}^n (Z_i - \hat{Z}_i)^2 = \frac{1}{4} \sum_{i=1}^n (Y_i - \hat{Y}_i)^2$
- ④  $\frac{\sum_{i=1}^n (\hat{Z}_i - \bar{Z})^2}{\sum_{i=1}^n (Z_i - \bar{Z})^2} = \frac{1}{4} \times \frac{\sum_{i=1}^n (\hat{Y}_i - \bar{Y})^2}{\sum_{i=1}^n (Y_i - \bar{Y})^2}$

문 9. 평균이 10이고 분산이 1인 정규모집단에서 얻은 임의표본(random sample)을  $X_1, X_2, X_3, X_4$ 라고 할 때, 옳은 것만을 모두 고른 것은?

- ㄱ.  $X_1 - 2X_2 + X_3$ 은 평균이 0이고 분산이 6인 정규분포를 따른다.
- ㄴ.  $(X_4 - 10)^2$ 은 자유도가 1인 카이제곱분포를 따른다.
- ㄷ.  $\left(\frac{X_1 - X_2}{X_3 - X_4}\right)^2$ 은 분자의 자유도가 1, 분모의 자유도가 1인  $F$ 분포를 따른다.

- ① ㄱ
- ② ㄱ, ㄴ
- ③ ㄴ, ㄷ
- ④ ㄱ, ㄴ, ㄷ

문 10. 표본의 크기가 10인 자료에 단순선형회귀모형  $Y_i = \beta_0 + \beta_1 X_i + \epsilon_i, i = 1, 2, \dots, 10$ 을 적용하여 얻은 분산분석표는 다음과 같다. 만일  $\sum_{i=1}^{10} (X_i - \bar{X})^2 = 100, \sum_{i=1}^{10} (Y_i - \bar{Y})^2 = 800$ , 추정된 회귀계수가  $\hat{\beta}_1 = 2$ 라면, 분산분석표에 들어갈 값에 대한 설명으로 옳은 것만을 모두 고른 것은? (단,  $\bar{X} = \frac{1}{10} \sum_{i=1}^{10} X_i, \bar{Y} = \frac{1}{10} \sum_{i=1}^{10} Y_i, \epsilon_i$ 는 서로 독립이며 평균이 0이고 분산이  $\sigma^2$ 인 정규분포를 따른다)

|    |     |     |      |     |
|----|-----|-----|------|-----|
| 요인 | 제곱합 | 자유도 | 평균제곱 | F-값 |
| 회귀 | A   |     | B    | C   |
| 잔차 |     | D   | E    |     |
| 계  |     |     |      |     |

- ㄱ. A의 값은 400이다.
- ㄴ. C의 값과 E의 값의 합은 58이다.
- ㄷ. A와 B의 값은 같다.
- ㄹ. C와 D의 값은 같지 않다.

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

문 11. 다음의 줄기-잎 그림(stem-and-leaf plot)에 대한 설명으로 옳은 것만을 모두 고른 것은? (단, 줄기는 십의 자릿수이고 잎은 일의 자릿수이다)

| <자료 A> |         | <자료 B> |           |
|--------|---------|--------|-----------|
| 줄기     | 잎       | 줄기     | 잎         |
| 1      | 1 3 5 7 | 1      | 1         |
| 2      | 1 3     | 2      | 3 5 7     |
| 3      | 1       | 3      | 1 3 5 7 9 |
| 4      | 1 3     | 4      | 3 5 7     |
| 5      | 1 3 5 7 | 5      | 1         |

- ㄱ. <자료 A>의 범위가 <자료 B>의 범위보다 크다.
- ㄴ. <자료 A>의 평균과 중앙값은 같다.
- ㄷ. <자료 A>와 <자료 B>의 중앙값은 같다.

- ① ㄱ
- ② ㄱ, ㄴ
- ③ ㄴ, ㄷ
- ④ ㄱ, ㄴ, ㄷ

문 12. 평균이 0이고 분산이 1인 확률변수  $X$ 에 대하여  $Y = 1 - 2X$ 일 때, 옳지 않은 것은?

- ①  $Y$ 의 평균은 1이다.
- ②  $Y$ 의 표준편차는 2이다.
- ③  $X$ 와  $Y$ 의 공분산은  $-2$ 이다.
- ④  $X$ 와  $Y$ 의 상관계수는 1이다.

문 13. 정책 A, B에 대한 성별 선호도를 알아보기 위해 남자 100명과 여자 100명을 무작위로 추출하여 조사한 결과가 다음 표와 같다. 남자와 여자가 정책 A를 선호하는 비율을 각각  $p_{1A}, p_{2A}$ 라고 할 때, 가설  $H_0: p_{1A} = p_{2A}$  대  $H_1: p_{1A} \neq p_{2A}$ 를 검정하기 위한 카이제곱검정통계량의 값과 자유도는?

|    | 정책 A | 정책 B |
|----|------|------|
| 남자 | 40   | 60   |
| 여자 | 50   | 50   |

- |   |     |
|---|-----|
| 값   | 자유도 |
| ① $\frac{(40-45)^2}{45} + \frac{(60-55)^2}{55} + \frac{(50-45)^2}{45} + \frac{(50-55)^2}{55}$ | 1   |
| ② $\frac{(40-45)^2}{45} + \frac{(60-55)^2}{55} + \frac{(50-45)^2}{45} + \frac{(50-55)^2}{55}$ | 2   |
| ③ $\frac{(40-50)^2}{50} + \frac{(60-50)^2}{50} + \frac{(50-50)^2}{50} + \frac{(50-50)^2}{50}$ | 1   |
| ④ $\frac{(40-50)^2}{50} + \frac{(60-50)^2}{50} + \frac{(50-50)^2}{50} + \frac{(50-50)^2}{50}$ | 2   |

- 문 14. 평균이  $\mu$ 이고 분산이  $\sigma^2$ 인 정규모집단에서 얻은 임의표본(random sample)을  $X_1, X_2, \dots, X_n$ 이라고 하자.  $\sigma^2$ 의 추정량을  $\hat{\sigma}^2 = \frac{1}{n} \sum_{i=1}^n (X_i - \bar{X})^2$ 이라고 할 때,  $\hat{\sigma}^2$ 의 성질로 옳은 것은? (단,  $\bar{X} = \frac{1}{n} \sum_{i=1}^n X_i$ )
- ①  $\hat{\sigma}^2$ 은 불편추정량이 아니며 일치추정량도 아니다.
  - ②  $\hat{\sigma}^2$ 은 불편추정량은 아니지만 일치추정량이다.
  - ③  $\hat{\sigma}^2$ 은 불편추정량이지만 일치추정량은 아니다.
  - ④  $\hat{\sigma}^2$ 은 불편추정량이며 일치추정량이다.

- 문 15. 실험 대상자 50명 중 25명을 무작위로 추출하여 치료방법 A를 적용하고, 나머지 25명은 치료방법 B를 적용하여 회복시간을 조사한 결과가 다음 표와 같다. 두 치료방법에 따른 환자의 회복시간은 각각 정규분포를 따르고 분산은 서로 같다고 가정하자. 치료방법 A, B의 평균회복시간을 각각  $\mu_A, \mu_B$ 라고 할 때,  $(\mu_A - \mu_B)$ 에 대한 95% 신뢰구간은? (단,  $t_\alpha(m)$ 은 자유도가  $m$ 인  $t$ 분포의 제  $(1-\alpha) \times 100$  백분위수를 나타낸다)

| 구분   | 치료방법 |    |
|------|------|----|
|      | A    | B  |
| 표본평균 | 10   | 15 |
| 표본분산 | 3    | 6  |

- ①  $-5 \pm t_{0.025}(48) \sqrt{\frac{9}{25}}$
- ②  $-5 \pm t_{0.025}(48) \sqrt{\frac{9}{50}}$
- ③  $-5 \pm t_{0.025}(24) \sqrt{\frac{9}{25}}$
- ④  $-5 \pm t_{0.025}(24) \sqrt{\frac{9}{50}}$

- 문 16. 가설  $H_0$ : “모집단의 분포가  $U(-1,1)$ 이다.” 대  $H_1$ : “모집단의 분포가  $U(0,5)$ 이다.”를 검정하기 위하여 모집단에서 한 개의 표본  $X$ 를 무작위로 추출하였다. ‘만약  $X$ 가 0.8보다 크면 가설  $H_0$ 을 기각하고, 그렇지 않으면  $H_0$ 을 기각하지 않는다.’고 할 때, 이 검정법의 제 1종의 오류를 범할 확률과 검정력(power)은? (단,  $U(a,b)$ 는 구간  $[a,b]$ 에서 정의된 균일분포를 따른다)

|   | 제 1종의 오류를 범할 확률 | 검정력  |
|---|-----------------|------|
| ① | 0.1             | 0.84 |
| ② | 0.1             | 0.9  |
| ③ | 0.16            | 0.84 |
| ④ | 0.16            | 0.9  |

- 문 17. 표본의 크기가 11인 자료에 단순선형회귀모형  $Y_i = \beta_0 + \beta_1 X_i + \epsilon_i$ ,  $i = 1, 2, \dots, 11$ 을 적용하여 추정된 회귀직선의 결정계수(coefficient of determination)가 0.7이라고 한다. 이 회귀직선의 유의성검정을 위한  $F$  검정통계량의 값은? (단,  $\epsilon_i$ 는 서로 독립이며 평균이 0이고 분산이  $\sigma^2$ 인 정규분포를 따른다)
- ① 12
  - ② 15
  - ③ 18
  - ④ 21

- 문 18. 육면 주사위를 240번 던져 나온 결과가 다음 표와 같다. 가설  $H_0$ : “주사위의 각각의 눈이 나올 확률은  $\frac{1}{6}$ 이다.” 대  $H_1$ : “ $H_0$ 가 아니다.”를 검정할 때,  $p$ -값( $p$ -value 또는 유의확률)은? (단,  $\chi_\alpha^2(m)$ 은 자유도가  $m$ 인 카이제곱분포의 제  $(1-\alpha) \times 100$  백분위수를 나타내며,  $\chi_{0.05}^2(5) = 11.07$ ,  $\chi_{0.025}^2(5) = 12.83$ ,  $\chi_{0.01}^2(5) = 15.09$ ,  $\chi_{0.05}^2(6) = 12.59$ ,  $\chi_{0.025}^2(6) = 14.45$ ,  $\chi_{0.01}^2(6) = 16.81$ 이다)

| 눈    | 1  | 2  | 3  | 4  | 5  | 6  | 계   |
|------|----|----|----|----|----|----|-----|
| 관측도수 | 50 | 48 | 30 | 50 | 32 | 30 | 240 |

- ① 0.01보다 작거나 같다.
- ② 0.01보다 크고 0.025보다 작거나 같다.
- ③ 0.025보다 크고 0.05보다 작거나 같다.
- ④ 0.05보다 크다.

- 문 19. 살을 빼려는 20명을 대상으로 네 가지 식이요법을 각각 5명에게 3개월간 적용하여 얻은 체중 변화량에 대한 분산분석표의 일부가 다음과 같을 때, 옳지 않은 것은? (단,  $F_\alpha(m,n)$ 은 분자의 자유도가  $m$ , 분모의 자유도가  $n$ 인  $F$ 분포의 제  $(1-\alpha) \times 100$  백분위수를 나타낸다)

| 요인 | 제곱합  | 자유도 | 평균제곱 | F-값 | $p$ -값 |
|----|------|-----|------|-----|--------|
| 처리 | 20.4 | A   |      | 1.7 | 0.21   |
| 잔차 |      |     | B    |     |        |
| 계  | 84.4 |     |      |     |        |

- ① A = 3
- ② B = 4
- ③  $F_{0.05}(3,16)$ 은 1.7보다 크다.
- ④ 확률변수  $X$ 는 분자의 자유도가 3, 분모의 자유도가 16인  $F$ 분포를 따를 때  $P(X > 2)$ 는 0.21보다 크다.

- 문 20. 평균이  $\mu$ 이고 분산이 16인 정규모집단에서 얻은 임의표본(random sample)을  $X_1, X_2, \dots, X_9$ 라고 하자. 확률변수  $V = \frac{3(\bar{X} - \mu)}{4}$ 와

$W = \frac{3(\bar{X} - \mu)}{S}$ 에 대한 설명으로 옳은 것만을 모두 고른 것은?

(단,  $\bar{X} = \frac{1}{9} \sum_{i=1}^9 X_i$ ,  $S = \sqrt{\frac{1}{8} \sum_{i=1}^9 (X_i - \bar{X})^2}$ )

|   |
|---|
| ㄱ. $V$ 의 분포는 0에 대해 대칭이다.                                |
| ㄴ. $W$ 의 분포는 0에 대해 대칭이다.                                |
| ㄷ. 1보다 큰 모든 실수 $c$ 에 대해 $P(V \geq c) < P(W \geq c)$ 이다. |

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