

# 통 계 학

문 1. 어떤 양적 자료에 대한 상자그림(box plot)으로부터 알 수 없는 통계량은?

- ① 범위(range)
- ② 중앙값(median)
- ③ 최빈값(mode)
- ④ 사분위수(quartile)

문 2. 두 이산확률변수  $X$ 와  $Y$ 의 결합확률분포가 다음과 같을 때,  $X$ 와  $Y$  사이의 상관계수는?

$X \backslash Y$	-1	0	1
-1	0	$\frac{1}{3}$	0
0	0	0	$\frac{1}{3}$
1	$\frac{1}{3}$	0	0

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

문 3. 어느 자동차 보험회사는 모든 고객을 사고위험군과 비사고위험군으로 분류하며, 30%의 고객이 사고위험군에 속한다고 한다. 사고위험군에 속한 고객 중에서 보험에 가입한 지 1년 이내에 보험료를 청구할 확률은 0.4, 비사고위험군에 속한 고객 중에서 보험에 가입한 지 1년 이내에 보험료를 청구할 확률은 0.2이다. 이 회사의 고객이 보험에 가입한 지 1년 이내에 보험료를 청구하였을 때, 이 고객이 사고위험군에 속할 확률은?

- ①  $\frac{6}{13}$
- ②  $\frac{7}{13}$
- ③  $\frac{1}{2}$
- ④  $\frac{7}{12}$

문 4. 어떤 지역의 백화점 A와 B를 이용하는 고객의 나이는 분산이 각각 48과 50인 정규분포를 따른다고 한다. 백화점 A를 이용하는 고객 중 12명, 백화점 B를 이용하는 고객 중 10명을 임의추출하였을 때 나이의 표본평균은 각각 40과 34이며, 두 표본은 서로 독립이다. 백화점 A를 이용하는 고객의 평균나이를  $\mu_A$ , 백화점 B를 이용하는 고객의 평균나이를  $\mu_B$ 라 할 때,  $\mu_A - \mu_B$ 에 대한 95% 신뢰구간은? (단,  $z_\alpha$ 와  $t_\alpha(k)$ 는 각각 표준정규분포와 자유도가  $k$ 인  $t$ 분포의  $(1-\alpha) \times 100$ 번째 백분위수를 나타낸다)

- ①  $6 \pm t_{0.025}(20) \times 3$
- ②  $6 \pm t_{0.025}(20) \times 7$
- ③  $6 \pm z_{0.025} \times 3$
- ④  $6 \pm z_{0.025} \times 7$

문 5. 평균이  $\mu$ 이고 분산이  $\sigma^2$ 인 정규모집단에서의 임의표본  $X_1, X_2, \dots, X_n$ 에 대하여  $\bar{X} = \frac{1}{n} \sum_{i=1}^n X_i$  이고  $S^2 = \frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X})^2$  일 때, 다음 중 옳은 것으로만 묶인 것은? (단, 표본의 크기  $n$ 은 1보다 크다)

- ㄱ.  $\bar{X}$ 와  $S^2$ 은 서로 독립이다.
- ㄴ.  $\bar{X}$ 는 평균이  $\mu$ 이고 분산이  $\sigma^2$ 인 정규분포를 따른다.
- ㄷ.  $S^2$ 은  $\sigma^2$ 의 불편추정량(unbiased estimator)이다.
- ㄹ. 확률변수  $\frac{nS^2}{\sigma^2}$ 은 자유도가  $n$ 인 카이제곱분포를 따른다.

- ① ㄱ, ㄴ
- ② ㄱ, ㄷ
- ③ ㄴ, ㄹ
- ④ ㄷ, ㄹ

문 6. 확률변수  $X$ 와  $Y$ 의 표준편차는 각각 2와 5이고 두 확률변수  $X$ 와  $Y$  사이의 상관계수가  $-\frac{1}{2}$ 일 때,  $X$ 와  $Y$  사이의 공분산은?

- ① -20
- ② -5
- ③  $-\frac{1}{5}$
- ④  $-\frac{1}{20}$

문 7. 설명변수(또는 독립변수)를  $X$ , 반응변수(또는 종속변수)를  $Y$ 로 하는 단순선형회귀모형  $Y_i = \beta_0 + \beta_1 X_i + \epsilon_i$ ,  $i = 1, \dots, n$ 에서 다음 통계량을 이용하여 구한 결정계수(coefficient of determination)는?

$$n = 5, \bar{X} = 1, \sum_{i=1}^n X_i^2 = 10, \bar{Y} = 2, \sum_{i=1}^n Y_i^2 = 30, \sum_{i=1}^n X_i Y_i = 5$$

- ① 0.5
- ② 0.6
- ③ 0.7
- ④ 0.8

문 8. 두 인자 A와 B는 각각 세 가지 수준을 가지며, 두 인자의 아홉 가지 수준조합별로 두 번씩 반복실험을 하여 반응변수에 미치는 효과를 알아보고자 한다. 이원배치 분산분석법(two-way analysis of variance)을 적용하여 얻은 분산분석표의 일부는 다음과 같다. 반응변수에 미치는 효과에 대한 설명으로 옳은 것은? (단,  $F_\alpha(m, n)$ 은 분자의 자유도가  $m$ , 분모의 자유도가  $n$ 인  $F$ 분포의  $(1-\alpha) \times 100$ 번째 백분위수를 나타내고,  $F_{0.05}(2, 9) = 4.26$ ,  $F_{0.05}(4, 9) = 3.63$ 이다)

요인	제곱합	자유도	평균제곱
인자 A	50	2	25
인자 B	90	2	45
교호작용	80	4	20
오차	90	9	10
합계	310	17	

- ① 인자 A의 효과는 유의수준 5%에서 유의하다.
- ② 인자 B의 효과는 유의수준 5%에서 유의하다.
- ③ 인자 A와 B 간 교호작용의 효과는 유의수준 5%에서 유의하다.
- ④ 인자 A와 B 간 교호작용의 효과는 유의수준 1%에서 유의하다.



문 18. 어떤 회사에서 세 종류 기계 A, B, C의 일일 평균 생산량이 차이가 있는지를 알아보기 위하여 일원배치 분산분석법(one-way analysis of variance)으로 실험한 결과가 다음과 같다.

	기계 A	기계 B	기계 C
일일 생산량	25	21	22
	20	20	20
	25	16	21
	26	15	

위의 자료를 이용하여 만든 아래 분산분석표의 일부에서 (가)와 (나)의 값으로 옳은 것은?

요인	제곱합	자유도	F 비
처리	72		(나)
오차	50	(가)	

- |     |                  |
|-----|------------------|
| (가) | (나)              |
| ① 8 | $\frac{144}{25}$ |
| ② 8 | $\frac{162}{25}$ |
| ③ 9 | $\frac{144}{25}$ |
| ④ 9 | $\frac{162}{25}$ |

문 19. 어떤 안전에 대한 의견을 알아보기 위하여 모집단을 몇 개의 그룹으로 나눈 후 각 그룹별로 임의로 표본을 추출하여 설문조사를 하고자 한다. 그룹을 나눌 때, 이 안전에 대하여 같은 그룹에 속하는 구성원들 간의 의견은 동질성이 높도록 하고 다른 그룹에 속하는 구성원들 간의 의견은 이질성이 높도록 한다. 이 때 사용한 표본추출방법은?

- ① 단순임의표집법(simple random sampling)
- ② 군집(집락)표집법(cluster sampling)
- ③ 계통표집법(systematic sampling)
- ④ 층화표집법(stratified sampling)

문 20. 서로 독립인 세 개의 확률변수  $X_i, i = 1, 2, 3$ 은 각각 평균이  $i$ 이고 분산이  $i^2$ 인 정규분포를 따른다고 할 때, 다음 중 옳지 않은 것은?

- ①  $\sum_{i=1}^3 \left( \frac{X_i - i}{i} \right)^2$ 은 자유도가 3인 카이제곱분포를 따른다.
- ②  $\frac{3(X_1 - 1)}{\sqrt{(X_3 - 3)^2}}$ 은 자유도가 1인  $t$  분포를 따른다.
- ③  $\frac{4(X_1 - 1)^2}{(X_2 - 2)^2}$ 은 분자의 자유도가 1, 분모의 자유도가 2인  $F$  분포를 따른다.
- ④  $(X_1 - 1)^2$ 은 자유도가 1인 카이제곱분포를 따른다.