

통 계 학

- 문 1. 다음은 A 지역에서 20일간 측정된 미세먼지의 양을 요약한 것이다. 이에 대한 설명으로 옳은 것은? (단, $100\alpha\%$ 절사평균은 관측값의 양쪽에서 각각 일정비율 α 의 극단값을 버린 나머지 관측값들의 평균이다)

표본크기	평균	표준편차	중앙값	최솟값	최댓값
20	50.0	58.9	20	10	180

- ① 변동계수는 1보다 작다.
 ② 평균을 중심으로 대칭인 분포이다.
 ③ 범위(range)는 180이다.
 ④ 5% 절사평균은 45.0이다.
- 문 2. 측정 단위의 영향을 받지 않는 통계량만을 모두 고르면?

- ㄱ. 표준편차(standard deviation)
 ㄴ. 변동계수(coefficient of variation)
 ㄷ. 산술평균(arithmetic mean)
 ㄹ. 중앙값(median)
 ㅁ. 피어슨 상관계수(Pearson correlation coefficient)
 ㅂ. 공분산(covariance)

- 문 3. 공정한 동전을 던지는 놀이를 하고 있다. 확률변수 X 를 앞면이 나올 때까지 동전을 던진 횟수라 하자. X 가 1이면 1,000원, 2이면 800원, 3이면 400원, 4이상이면 0원을 상금으로 받는다고 할 때, 참가비가 800원이라면 이 놀이에서 참가자가 얻는 이익의 기대 금액은?

- ① -100원 ② -50원
 ③ 0원 ④ 100원
- 문 4. 확률변수 X 가 이항분포 $B(16, p)$ 를 따른다. 분산 $Var(X)$ 가 최대일 때의 평균 $E(X)$ 는?
- ① 4 ② 6
 ③ 8 ④ 10

- 문 5. A 대학교 재학생 중 40%가 남학생, 60%가 여학생이다. 남학생 중 기숙사 입주자의 비율을 p_1 , 여학생 중 기숙사 입주자의 비율을 p_2 라 할 때, $\frac{p_1}{p_2}=3$ 이다. 기숙사 입주자 중 남학생의 비율을 a_1 , 기숙사 입주자 중 여학생의 비율을 a_2 라 할 때 $\frac{a_1}{a_2}$ 의 값은?

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

- 문 6. 연속형 확률변수 X 의 확률밀도함수(probability density function)가 다음과 같을 때, 옳은 것만을 모두 고르면?

$$f(x) = \begin{cases} x + \frac{1}{2}, & 0 \leq x \leq a \\ 0, & \text{그 밖의 경우} \end{cases}$$

ㄱ. a 의 값은 2이다.

ㄴ. $f(2) = \frac{5}{2}$

ㄷ. $P(0 \leq X \leq \frac{1}{2}) = \frac{3}{8}$

- ① ㄷ ② ㄱ, ㄴ
 ③ ㄴ, ㄷ ④ ㄱ, ㄴ, ㄷ

- 문 7. 지난달 A 드라마 시청률이 50%였다. 이번 달의 A 드라마 시청률이 지난달과 다른지를 이항검정법으로 검정하기 위하여 10세대를 랜덤 추출하여 이번 달 A 드라마를 시청한 세대수(X)를 조사하고자 한다. 유의수준 10%에서 귀무가설(H_0)의 기각역은? (단, p 는 이번 달의 A 드라마 시청률이다)

$$H_0: p = \frac{1}{2}, H_1: p \neq \frac{1}{2}$$

- ① $X=0$ 또는 $X=10$
 ② $X \leq 1$ 또는 $X \geq 9$
 ③ $X \leq 2$ 또는 $X \geq 8$
 ④ $X \leq 3$ 또는 $X \geq 7$

- 문 8. 표준편차가 10인 정규모집단에서 모평균(μ)에 대한 귀무가설 $H_0: \mu = 10$ 대 대립가설 $H_1: \mu = 12$ 를 검정하고자 한다. 크기 100의 표본을 랜덤 추출하여 표본평균 \bar{X} 를 구한 후 기각역으로 $R: \bar{X} \geq 11.5$ 를 사용할 때, 제2종 오류의 확률은? (단, Z 는 표준정규분포를 따르는 확률변수를 나타낸다)

- ① $P(Z < -0.05)$ ② $P(Z < -0.5)$
 ③ $P(Z > 0.15)$ ④ $P(Z > 1.5)$

- 문 9. T 는 자유도가 k 인 t 분포를 따르는 확률변수를 나타내고, $t_\alpha(k)$ 는 자유도가 k 인 t 분포의 제100(1- α)백분위수를 나타낼 때, 옳은 것만을 모두 고르면? (단, $k > 2$, $0 < \alpha < 1$ 이다)

ㄱ. T 의 평균은 0이다.

ㄴ. α 가 고정되어 있을 때, k 가 증가할수록 $t_\alpha(k)$ 는 감소한다.

ㄷ. k 가 고정되어 있을 때, α 가 증가할수록 $t_\alpha(k)$ 는 감소한다.

- ① ㄴ ② ㄱ, ㄴ
 ③ ㄱ, ㄷ ④ ㄱ, ㄴ, ㄷ

문 10. 단순선형회귀모형 $Y_i = \beta_0 + \beta_1 x_i + \epsilon_i$, $i = 1, 2, \dots, 10$ 을 최소
제곱법으로 적합시킨 회귀식이 $\hat{y}_i = 1 + \frac{1}{3}x_i$ 이고 $\sum_{i=1}^{10} \hat{y}_i = 70$ 일 때,
 $E[Y|x]$ 의 95% 신뢰구간의 길이가 최소가 되는 x 의 값은?
(단, ϵ_i 는 $N(0, \sigma^2)$ 을 따르며 서로 독립이다)

- ① 12 ② 18
③ 21 ④ 24

문 11. 어느 병원에 내원한 호흡기 질환자 중 200명을 랜덤 추출하여
계절별로 정리하면 다음과 같다. ‘호흡기 질환자의 계절별 비율은
동일하다’는 귀무가설을 검정하기 위해 적합한 검정법과 유의수준
5%에서의 검정 결과를 바르게 연결한 것은? (단, 자유도가
 k_1, k_2 인 F 분포의 제100(1- α)백분위수를 $F_\alpha(k_1, k_2)$ 로 나타낼 때
 $F_{0.05}(1, 3) = 10.13$, $F_{0.05}(1, 4) = 7.71$ 이며, 자유도가 k 인 χ^2 분포의
제100(1- α)백분위수를 $\chi_\alpha^2(k)$ 로 나타낼 때 $\chi_{0.05}^2(3) = 7.81$,
 $\chi_{0.05}^2(4) = 9.49$ 이다)

계절	봄	여름	가을	겨울
환자 수	40	65	40	55

검정법

검정 결과

- ① F 귀무가설을 기각하지 않음
② F 귀무가설을 기각함
③ χ^2 귀무가설을 기각하지 않음
④ χ^2 귀무가설을 기각함

문 12. 단순선형회귀모형 $Y_i = \beta_0 + \beta_1 x_i + \epsilon_i$, $i = 1, 2, \dots, n$ 을 최소
제곱법으로 적합시켰을 때 결정계수와 같은 값을 갖는 것만을
모두 고르면? (단, ϵ_i 는 $N(0, \sigma^2)$ 을 따르고 서로 독립이며,
 $\hat{\beta}_1$ 는 β_1 에 대한 추정량, SSE 는 잔차제곱합, $S_{xx} = \sum_{i=1}^n (x_i - \bar{x})^2$,
 $S_{yy} = \sum_{i=1}^n (y_i - \bar{y})^2$, $S_{xy} = \sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})$, $\bar{x} = \frac{1}{n} \sum_{i=1}^n x_i$,
 $\bar{y} = \frac{1}{n} \sum_{i=1}^n y_i$ 이다)

$$\neg. 1 - \frac{SSE}{S_{yy}}$$

$$\neg. \frac{\hat{\beta}_1^2 S_{xx}}{S_{yy}}$$

$$\neg. \frac{(S_{xy})^2}{S_{xx} S_{yy}}$$

- ① \neg ② \neg, \neg
③ \neg, \neg ④ \neg, \neg, \neg

문 13. 다음은 서로 다른 세 가지 비료에 따른 평균수확량의 차이가
있는지 알아보기 위한 분산분석표이다. 이에 대한 설명으로 옳은
것만을 모두 고르면? (단, 자유도가 k_1, k_2 인 F 분포의
제100(1- α)백분위수를 $F_\alpha(k_1, k_2)$ 로 나타낼 때,
 $F_{0.05}(2, 27) = 3.35$, $F_{0.05}(3, 26) = 2.98$ 이다)

요인	제곱합	자유도	평균제곱	F -값
처리	17.956	()	8.978	10.125
오차	23.941	()	0.887	
전체	41.897	29		

- ㄱ. 유의수준 5%에서 세 가지 비료에 따른 평균수확량은
모두 다르다.
ㄴ. $F_{m,n}$ 이 자유도가 m, n 인 F 분포를 따르는 확률변수일 때
유의확률은 $P(F_{2,27} > 10.125)$ 이다.
ㄷ. 각 처리의 반복수가 모두 같다면 각 처리에서 9개씩
조사한 것이다.

- ① \neg
② \neg
③ \neg, \neg
④ \neg, \neg, \neg

문 14. 다음은 인자 A, B의 수준수가 각각 3이고, 각 처리에서 반복수가
3인 이원배치법 분산분석표의 일부이다. 이에 대한 설명으로 옳은
것은? (단, 자유도가 k_1, k_2 인 F 분포의 제100(1- α)백분위수를
 $F_\alpha(k_1, k_2)$ 로 나타낼 때 $F_{0.05}(2, 22) = 3.44$, $F_{0.05}(3, 20) = 3.10$ 이며,
인자 A, B의 교호작용은 고려하지 않는다)

요인	제곱합	자유도	평균제곱	F -값
인자 A	100			
인자 B	20			
오차				
계	230			

- ① 인자 A의 F -값이 인자 B의 F -값보다 작다.
② 인자 B의 유의확률은 0.05보다 작다.
③ 오차항의 자유도는 20이다.
④ 오차평균제곱은 5이다.

문 15. 다중선형회귀모형 $Y_i = \beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + \epsilon_i$, $i = 1, 2, \dots, n$ 에서
모수 $\beta_0, \beta_1, \beta_2$ 를 추정하기 위해 사용할 수 있는 방법은? (단,
오차항 ϵ_i 는 서로 독립이며 평균 0, 분산 σ^2 인 확률변수이다)

- ① 최소제곱법(least squares method)
② 최대우도법(maximum likelihood estimation)
③ 우도비 검정법(likelihood ratio test)
④ 분산분석법(ANOVA)

문 16. X_1 과 X_2 는 서로 독립이고 평균이 10, 분산이 1인 정규분포를 따르는 확률변수이다. Y 가 X_1 과 X_2 중 최댓값일 때, Y 가 10보다 클 확률은?

- ① 0.5 ② 0.625
③ 0.75 ④ 0.875

문 17. 과목 A의 점수는 $N(50, 5^2)$ 을 따르고, 과목 B의 점수는 $N(60, 10^2)$ 을 따른다. 갑돌이의 과목 A, B 점수는 각각 45점, 80점이고, 갑순이의 과목 A, B 점수는 각각 50점, 70점이다. 다음 설명 중 옳은 것만을 모두 고르면? (단, Z 가 $N(0, 1)$ 을 따를 때 $P(Z \leq 1) = 0.84$, $P(Z \leq 2) = 0.98$ 이고, 어떤 사람의 과목 성적이 해당 과목의 점수 분포에서 하위 $p\%$ 지점에 위치하면 이 사람의 해당 과목 백분위점수는 p 라고 정의한다)

ㄱ. 두 과목 원점수의 합은 갑돌이가 더 높다.

나. 두 과목 표준화점수(z-score)의 합은 갑돌이와 갑순이가 같다.

ㄷ. 두 과목 백분위점수의 합은 갑순이가 더 높다.

- ① \neg, \perp
② \neg, \sqsubset
③ \perp, \sqsubset
④ \neg, \perp, \sqsubset

문 18. 두 정규모집단에서 각각 랜덤 추출한 서로 독립인 표본 1과 표본 2의 통계값이 다음과 같다. 이에 대한 설명으로 옳은 것만을 모두 고르면? (단, 자유도가 k_1, k_2 인 F 분포의 제 $100(1-\alpha)$ 백분위수를 $F_\alpha(k_1, k_2)$ 로 나타낼 때, $F_{0.025}(9, 11) = 3.59$, $F_{0.025}(11, 9) = 3.91$, $F_{0.05}(9, 11) = 2.90$, $F_{0.05}(11, 9) = 3.10$ 이다)

구분	표본크기	평균	표준편차
표본 1	10	10	3
표본 2	12	15	2

7. 유의수준 5%에서 두 집단의 모분산은 다르다고 할 수 없다.

나. 두 집단의 모분산이 같다는 가정하에, 두 표본의 분산을 사용하여 공통분산의 불편추정값을 구하면 6.25이다.

다. 두 집단의 모분산이 같다는 가정하에, 두 집단의 모평균이 같다는 귀무가설과 같지 않다는 대립가설을 검정하기

위해 검정통계량의 값 $\frac{5}{2.5 \times \sqrt{\frac{1}{10} + \frac{1}{12}}}$ 를 사용할 수

있다.

- ① \neg, \bot
- ② \neg, \top
- ③ \bot, \top
- ④ \neg, \bot, \top

문 19. 단순선형회귀모형 $Y_i = \beta_0 + \beta_1 X_i + \epsilon_i$, $i = 1, 2, \dots, 11$ 에 최소 제곱법을 적용하여 얻은 분산분석표의 일부가 다음과 같을 때, 옳지 않은 것은? (단, ϵ_i 는 $N(0, \sigma^2)$ 을 따르고 서로 독립이며, 자유도가 k_1, k_2 인 F 분포의 제100(1- α) 백분위수를 $F_\alpha(k_1, k_2)$ 로 나타낼 때 $F_{0.01}(1, 9) = 10.56$ 이다)

요인	제곱합	자유도	평균제곱	F -값
회귀	75			
오차(잔차)	45			
합계	120			

- ① X, Y 의 표본상관계수는 $\sqrt{0.625}$ 이다.
- ② 오차항 분산의 불편추정값은 5이다.
- ③ F -값은 15이다.
- ④ 유의수준 5%에서 회귀모형이 유의하다.

문 20. 중위소득 이상인 가구와 미만인 가구를 각각 100개씩 랜덤 추출하여
특정 정책에 대한 가구주의 의견을 조사한 결과 다음과 같은
차별 교차표를 얻었다.

구분	찬성	반대	합
중위소득 미만	60	40	100
중위소득 이상	40	60	100
합	100	100	200

소득 수준에 따라 찬성률의 차이가 없다는 귀무가설과 차이가 있다는 대립가설을 검정하기 위해 다음 두 가지 통계값을 사용하고자 한다.

$$a = \frac{0.6 - 0.4}{\sqrt{0.5 \times 0.5 \times (\frac{1}{100} + \frac{1}{100})}}$$
$$b = \frac{(60 - 50)^2}{50} + \frac{(40 - 50)^2}{50} + \frac{(60 - 50)^2}{50} + \frac{(40 - 50)^2}{50}$$

이에 대한 설명으로 옳은 것만을 모두 고르면?

7. a 를 사용하여 정규검정법으로 검정할 수 있다.

4. b 를 사용하여 χ^2 검정법으로 검정할 수 있다.

ㄷ. a 를 사용한 검정의 유의확률이 b 를 사용한 검정의
유의확률보다 작다.

- ① \neg
- ② \neg, \neg
- ③ \neg, \sqcup
- ④ \neg, \neg, \sqcup