

통 계 학

문 1. 다음은 어느 회사의 방문 고객 수에 대한 자료이다. 산술평균, 중앙값, 최빈값을 모두 더한 값은?

1, 2, 3, 3, 2, 3, 4, 6

- ① 7
- ② 8
- ③ 9
- ④ 10

문 2. 다음 사례에 해당하는 표본추출 방법은?

어느 여론조사 기관이 전국의 유권자 중 무작위로 뽑힌 1,000명에게 차기 대통령 후보에 대한 선호도를 물었다. 이때, 표본의 40%가 A 후보를, 35%가 B 후보를, 15%가 C 후보를 지지하는 결과를 얻었다.

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

문 3. 확률질량함수 또는 확률밀도함수가 될 수 없는 것은?

- ① $f(x) = \frac{1}{10}, (0 < x \leq 10 \text{인 정수})$
- ② $f(x) = \left(\frac{1}{2}\right)^x, (x \text{는 자연수})$
- ③ $f(x) = e^{-x}, (x > 0)$
- ④ $f(x) = \frac{1}{3}(3-2x), (0 \leq x \leq 1)$

문 4. 다음은 표본으로 조사된 5명의 양팔 길이(X)와 키(Y)에 대한

자료 $(x_i, y_i), i = 1, \dots, 5$ 이다. $\sum_{i=1}^5 (x_i - \bar{x})(y_i - \bar{y}) = 23.6$ 일 때,

양팔 길이와 키의 표본상관계수는? (단, 단위는 cm이다)

구분	1	2	3	4	5	평균	표준편차
양팔 길이(X)	168	165	168	169	162	166.4	2.88
키(Y)	169	165	167	167	163	166.2	2.28

- ① $\frac{4 \times 2.88 \times 2.28}{23.6}$
- ② $\frac{23.6}{4 \times 2.88 \times 2.28}$
- ③ $\frac{5 \times 2.88 \times 2.28}{23.6}$
- ④ $\frac{23.6}{5 \times 2.88 \times 2.28}$

문 5. 어떤 확률변수 X 의 확률분포표가 다음과 같다. 이 모집단에서 복원추출한 크기가 2인 표본으로부터 얻은 표본평균을 $\bar{X}$ 라고 하자. 표본평균 $\bar{X}$ 의 기댓값이 $E(\bar{X}) = \frac{1}{4}$ 일 때, 표본평균 $\bar{X}$ 의 분산 $\text{Var}(\bar{X})$ 은? (단, a 는 상수)

X	-1	0	1	합계
$P(X=x)$	$\frac{1}{2}-a$	a	$\frac{1}{2}$	1

- ① $\frac{9}{32}$
- ② $\frac{11}{32}$
- ③ $\frac{9}{16}$
- ④ $\frac{11}{16}$

문 6. $c(c \geq 2)$ 개의 범주로 나누어지는 $r(r \geq 2)$ 개의 다항모집단에서 $i(i = 1, \dots, r)$ 번째 모집단의 각 범주에 대한 모비율은 $p_{i1}, p_{i2}, \dots, p_{ic}$ 이고, 관측값은 $O_{i1}, O_{i2}, \dots, O_{ic}$ 이고, i 번째 모집단의 표본의 크기는 n_i 이고 $n = \sum_{i=1}^r n_i$ 이며, $j = 1, \dots, c$ 에 대해 $O_{.j} = \sum_{i=1}^r O_{ij}$ 이다. $r \times c$ 분할표의 동질성검정에 대한 설명으로 옳지 않은 것은?

- ① 모든 $j(j = 1, \dots, c)$ 에 대해 $\sum_{i=1}^r p_{ij} = 1$ 이다.
- ② 귀무가설은 각 $j(j = 1, \dots, c)$ 에 대해 $p_{1j} = p_{2j} = \dots = p_{rj}$ 이다.
- ③ 귀무가설하에서 $j(j = 1, \dots, c)$ 에 속할 공통인 각 범주의 모비율의 추정값은 $\frac{O_{.j}}{n}$ 이다.
- ④ 귀무가설하에서 i 번째 모집단의 j 번째 범주의 추정 기대도수는 $\frac{n_i O_{.j}}{n}$ 이다.

문 7. 질병에 대한 치료효과를 검정하기 위해 환자 10명의 치료 전과 치료 후의 검사치가 다음 표와 같다. 여기에서 $X_1, X_2, \dots, X_{10}$ 은 서로 독립이며 정규분포 $N(\mu_X, \sigma_X^2)$ 를, $Y_1, Y_2, \dots, Y_{10}$ 은 서로 독립이며 정규분포 $N(\mu_Y, \sigma_Y^2)$ 를 따른다. 치료 전후 검사치의 차의 모평균 μ_D 에 대해 귀무가설 $H_0: \mu_D = 0$ 대 대립가설 $H_1: \mu_D > 0$ 을 유의수준 α 에서 검정하고자 한다. 이에 대한 설명으로 옳은 것은? (단, $t_\alpha(n)$ 는 자유도 n 인 t 분포의 제 $100 \times (1 - \alpha)$ 백분위수이다)

환자	1	2	3	...	10	평균	분산
시기							
치료 전	X_1	X_2	X_3	...	X_{10}	$\bar{X}$	S_X^2
치료 후	Y_1	Y_2	Y_3	...	Y_{10}	$\bar{Y}$	S_Y^2
차이($D_i = X_i - Y_i$)	D_1	D_2	D_3	...	D_{10}	$\bar{D}$	S_D^2

- ① 검정통계량 $\frac{\bar{D}}{S_D/\sqrt{10}}$ 이 $-t_\alpha(9)$ 보다 작으면 귀무가설을 기각한다.
- ② 검정통계량 $\frac{\bar{D}}{S_D/\sqrt{10}}$ 이 $t_\alpha(9)$ 보다 크면 귀무가설을 기각한다.
- ③ 검정통계량 $\frac{\bar{D}}{S_D/\sqrt{10}}$ 이 $-t_\alpha(10)$ 보다 작으면 귀무가설을 기각한다.
- ④ 검정통계량 $\frac{\bar{D}}{S_D/\sqrt{10}}$ 이 $t_\alpha(10)$ 보다 크면 귀무가설을 기각한다.

문 8. 다음은 어떤 자료를 다중선형회귀모형 $Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \epsilon$ 에 적합하여 얻은 결과의 일부이다. 이에 대한 설명으로 옳지 않은 것은?
(단, 오차항 ϵ 은 평균이 0이고 분산이 σ^2 인 정규분포를 따른다)

모형	비표준화 계수	표준화 계수	...	유의확률
상수	1.793		...	
X_1	1.971	0.816	...	0.000
X_2	-1.290	-1.337	...	0.000

- ① Y 에 X_2 가 X_1 보다 더 크게 영향을 미친다.
- ② X_1, X_2 의 회귀계수는 유의수준 5%에서 모두 유의하다.
- ③ X_1 이 고정되어 있을 때 X_2 가 증가하면 Y 는 감소하는 경향이 있다.
- ④ X_2 가 고정되어 있을 때 X_1 이 1단위 증가하면 Y 의 추정값이 0.816배 증가한다.

문 9. $X_1, \dots, X_n$ 은 확률밀도함수 $f(x; \mu) = \frac{1}{\mu} e^{-\frac{x}{\mu}}$, ($\mu > 0, x > 0$)을 따르는 확률표본이다. 모수 μ 의 최대가능도추정량(maximum likelihood estimator) $\hat{\mu}$ 에 대한 설명으로 옳지 않은 것은? (단, log는 자연로그이다)

- ① 가능도함수(likelihood function)를 대수변환한 로그가능도함수는 $l(\mu) = -n \log \mu - \frac{1}{\mu} \sum_{i=1}^n x_i$ 이다.
- ② 로그가능도함수를 μ 에 대해 한 번 미분해서 0으로 놓고, 이 방정식의 해를 구한다.
- ③ 최대가능도추정량은 $\hat{\mu} = \frac{1}{n} \sum_{i=1}^n X_i$ 이다.
- ④ 최대가능도추정량은 편의추정량(biased estimator)이다.

문 10. X_1, X_2, X_3, X_4, X_5 는 서로 독립이며, 구간 $(0,1)$ 에서 균일분포 (uniform distribution)를 따르고, $X_{(1)} \leq X_{(2)} \leq X_{(3)} \leq X_{(4)} \leq X_{(5)}$ 는 X_1, X_2, X_3, X_4, X_5 의 순서통계량이라고 할 때, 기댓값 $E[X_{(4)}]$ 는?

- $$\begin{array}{ll} \textcircled{1} & \frac{1}{6} \\ \textcircled{2} & \frac{1}{3} \\ \textcircled{3} & \frac{2}{3} \\ \textcircled{4} & \frac{5}{6} \end{array}$$

문 11. 전체 인구의 5%가 어떤 질병을 가지고 있다고 한다. 이 질병을 진단하는 검사 방법이 있는데, 이 검사 방법은 이 질병을 실제로 가지고 있을 때 검사 결과가 양성(positive) 반응을 나타낼 확률이 0.95, 음성(negative) 반응을 나타낼 확률이 0.05이고, 이 질병을 실제로 가지고 있지 않을 경우 검사 결과가 양성 반응을 나타낼 확률이 0.1, 음성 반응을 나타낼 확률이 0.9이다. 임의로 한 사람을 선택하여 이 검사를 한 결과가 양성 반응으로 나타났다면, 이 사람이 실제로 이 질병을 가지고 있을 확률은?

- $$\begin{array}{ll} \textcircled{1} & \frac{1}{2} \\ \textcircled{2} & \frac{1}{3} \\ \textcircled{3} & \frac{1}{4} \\ \textcircled{4} & \frac{1}{5} \end{array}$$

문 12. 여론조사 회사에서 어떤 사항에 대한 국민의 찬성 비율 p 를 알아보기 위해 전화 설문조사를 무작위로 실시하려고 한다. 96% 신뢰수준에서 모든 p 에 대한 추정량의 오차한계를 0.05이내로 하는 최소 표본의 크기 n 을 구하는 부등식은? (단, z_α 는 표준정규분포의 제 $100 \times (1 - \alpha)$ 백분위수를 나타내고, $z_{0.04} = 1.75$, $z_{0.02} = 2.05$ 이고, 모집단의 크기는 충분히 크다)

- $$\begin{aligned} \textcircled{1} \quad n &\geq \left(\frac{1.75}{0.05}\right)^2 \frac{1}{2} \\ \textcircled{2} \quad n &\geq \left(\frac{1.75}{0.05}\right)^2 \frac{1}{4} \\ \textcircled{3} \quad n &\geq \left(\frac{2.05}{0.05}\right)^2 \frac{1}{2} \\ \textcircled{4} \quad n &\geq \left(\frac{2.05}{0.05}\right)^2 \frac{1}{4} \end{aligned}$$

문 13. 어느 가게에 월요일부터 금요일까지 방문하는 손님 수는 다음 표와 같다. 손님 수가 요일에 따라 다른지를 검정하기 위한 χ^2 검정통계량의 값과 유의수준 5%에서 검정결과를 바르게 연결한 것은? (단, $\chi^2_{\alpha}(k)$ 는 자유도가 k 인 χ^2 분포의 제 $100 \times (1 - \alpha)$ 백분위수를 나타내고, $\chi^2_{0.05}(4) = 9.49$, $\chi^2_{0.05}(5) = 11.07$ 이다)

요일	월	화	수	목	금	합계
손님 수(명)	25	25	10	15	25	100

χ^2 검정통계량의 값		검정결과
①	10	귀무가설을 기각할 수 없음
②	10	귀무가설을 기각함
③	11	귀무가설을 기각할 수 없음
④	11	귀무가설을 기각함

문 14. 다음은 처리 수준의 수가 5인 공정에서 처리에 따른 반응값의 모평균이 차이가 있는지를 검정하기 위한 일원배치 분산분석법을 적용한 결과의 일부이다. 검정에 대한 유의확률(significance probability) p 의 범위는? (단, $F_{\alpha}(k_1, k_2)$ 는 분자의 자유도가 k_1 이고 분모의 자유도가 k_2 인 F 분포의 제 $100 \times (1 - \alpha)$ 백분위수를 나타내고, $F_{0.01}(4, 17) = 4.67$, $F_{0.05}(4, 17) = 2.96$, $F_{0.1}(4, 17) = 2.31$ 이다)

요인	제공합	자유도	평균제공	F -값
처리	220	4		
잔차				
전체	390	21		

- ① $0 < p < 0.01$
- ② $0.01 < p < 0.05$
- ③ $0.05 < p < 0.1$
- ④ $0.1 < p < 1$

문 15. 관측된 자료 (x_i, y_i) , $i = 1, \dots, 21$ 에서 얻은 통계량의 값이

$\sum_{i=1}^{21} (x_i - \bar{x})^2 = 81$, $\sum_{i=1}^{21} (y_i - \bar{y})^2 = 100$ 이다. 이 자료를 단순선형

회귀모형 $Y_i = \beta_0 + \beta_1 X_i + \epsilon_i$ 에 적합하여 얻은 기울기에 대한
최소제곱추정값이 $\hat{\beta}_1 = -1$ 일 때, 이에 대한 설명으로 옳은 것은?

(단, 오차항 ϵ_i 는 서로 독립이며 정규분포 $N(0, \sigma^2)$ 을 따른다)

- ① X 가 증가할 때 Y 가 증가하는 경향이 있다.
- ② 설명변수와 반응변수의 상관계수는 0.9이다.
- ③ 회귀모형의 결정계수는 0.81이다.
- ④ 오차항의 분산 σ^2 에 대한 추정값인 평균제곱오차는 10이다.

문 16. 처리 수준의 수가 4인 인자 A 와 처리 수준의 수가 2인 인자 B 에 대한 반복이 없는 이원배치법의 실험에서 다음과 같은 자료를 얻었다. 분산분석 결과에서 인자 A 와 인자 B 가 모두 유의수준 5%에서 유의한 것으로 판명되었다. A_2 수준과 B_2 수준을 조합한 실험 조건에서 모평균의 추정값은?

인자 B \ 인자 A	A_1	A_2	A_3	A_4	합계
B_1	80	38	98	56	272
B_2	42	22	70	42	176
합계	122	60	168	98	448

- ① 18 ② 22
③ 30 ④ 44

문 17. 숫자 1, 2, 3, 4가 표시된 공정한 정사면체 A , B 두 개를 던져서 나오는 눈의 수를 각각 a , b 라고 하고, 확률변수 $X = \left\lfloor \frac{b}{a} \right\rfloor$ 라고 할 때, 확률변수 $Y = 16X$ 의 기댓값은? (단, $[x]$ 는 x 보다 크지 않는 최대 정수를 나타낸다)

- ① 15 ② 16
③ 17 ④ 18

문 18. $X_1, X_2, \dots, X_{10}$ 은 평균이 0, 분산이 4인 정규모집단으로부터 추출한 확률표본이고, $Y_1, Y_2, \dots, Y_{20}$ 은 평균이 0, 분산이 8인 정규모집단으로부터 추출한 확률표본이다. 두 모집단이 독립일 때,

표본평균 $\bar{X} = \frac{1}{10} \sum_{i=1}^{10} X_i$ 와 $\bar{Y} = \frac{1}{20} \sum_{i=1}^{20} Y_i$ 를 이용한 통계량

$$\left(\frac{\bar{X} + \bar{Y}}{\bar{X} - \bar{Y}} \right)^2$$
의 확률분포는?

- ① 자유도가 1인 카이제곱분포
- ② 자유도가 10인 카이제곱분포
- ③ 분자의 자유도가 1이고 분모의 자유도가 1인 F 분포
- ④ 분자의 자유도가 10이고 분모의 자유도가 20인 F 분포

문 19. 어느 자료에서 변수 X 와 Y 를 평균이 0, 표준편차가 1이 되도록 각각 표준화한 변수를 X^* 와 Y^* 라 하자. 설명변수 X^* 와 반응변수 Y^* 를 절편이 있는 단순선형회귀모형에 적용하여 얻은 추정회귀식과 분산분석표에 대한 설명으로 옳은 것만을 모두 고르면?

- ㄱ. 추정회귀식의 절편은 0이다.
- ㄴ. 분산분석표에서 회귀제곱합과 오차제곱합의 합은 회귀자유도와 오차자유도의 합과 같다.
- ㄷ. X 와 Y 가 각각 설명변수와 반응변수인 절편이 있는 단순선형회귀모형에 적용하여 얻은 분산분석표의 F -값과 X^* 와 Y^* 를 이용한 분산분석표의 F -값은 같다.

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

문 20. 다음은 어느 제품에 포함된 두 인자 A, B 의 수준에 따라 제품의 품질이 어떤 영향을 받는지 알아보기 위해 반복이 있는 이원배치 모형을 적용하여 얻은 분산분석표의 일부이다.

요인	제공합	자유도	평균제공	F -값
인자 A	16	1		
인자 B	14	1		
교호작용 $A \times B$	2			
오차	8			
전체		7		

여기서 교호작용의 효과가 유의수준 5%에서 유의하지 않아 교호작용을 포함하지 않는 이원배치 모형을 적용하여 재분석을 실시하였다. 다음은 재분석하여 얻은 분산분석표의 일부이다.

요인	제공합	자유도	평균제공	F -값
인자 A	16	1		
인자 B	14	1		
오차				
전체				

재분석 결과에 대한 설명으로 옳지 않은 것은? (단, $F_\alpha(k_1, k_2)$ 는 분자의 자유도가 k_1 이고 분모의 자유도가 k_2 인 F 분포의 제 $100 \times (1 - \alpha)$ 백분위수를 나타내고, $F_{0.05}(1, 4) = 7.71$, $F_{0.05}(1, 5) = 6.61$ 이다)

- ① 유의수준 5%에서 인자 A 의 효과가 있다고 할 수 있다.
- ② 유의수준 5%에서 인자 B 의 효과가 있다고 할 수 있다.
- ③ 오차항의 자유도는 4이다.
- ④ 평균제곱오차는 2이다.