

통계학개론

문 1. 다음 자료는 어느 고등학교 A, B반의 학생이 하루에 섭취하는 물의 양을 측정 한 것이다. A반은 리터(L) 단위로 측정하고, B반은 밀리리터(mL) 단위로 측정하였다. A, B반 학생이 섭취하는 물의 양에 대한 산포를 비교하고자 할 때 가장 적합한 측도는?

A반(L)	1.0	1.2	0.5	0.8	2.0	1.5	1.4	1.2	1.0	0.9
B반(mL)	500	700	450	600	1,000	300	500	500	1,500	400

- ① 표준편차
- ② 사분위수 범위
- ③ 제50백분위수
- ④ 변동계수

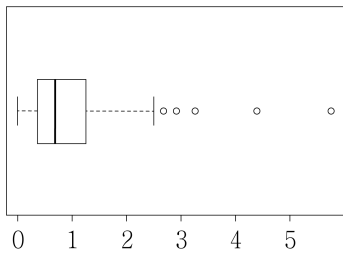
문 2. 두 사건 A, B에 대하여 $P(A) = \frac{1}{3}$, $P(A|B) = \frac{1}{9}$,

$P(A \cap B) = \frac{1}{12}$ 일 때, 옳은 것만을 모두 고르면?

- ㄱ. 사건 A와 B는 서로 배반이다.
- ㄴ. 사건 A와 B는 서로 독립이다.
- ㄷ. $P(A \cup B) = 1$ 이다.

- ① ㄴ
- ② ㄷ
- ③ ㄱ, ㄴ
- ④ ㄴ, ㄷ

문 3. 다음 상자그림에 대한 설명으로 옳지 않은 것은?



- ① 자료의 분포는 오른쪽으로 긴 꼬리를 가지는 비대칭 형태이다.
- ② 자료의 평균이 중앙값보다 작다.
- ③ 자료의 중앙값은 1보다 작다.
- ④ 자료의 왜도(skewness)는 정규분포의 왜도보다 크다.

문 4. 두 확률변수의 상관계수에 대한 설명으로 옳지 않은 것은?

- ① 상관계수는 항상 -1과 1 사이의 값이다.
- ② 공분산이 0이면 상관계수도 0이다.
- ③ 상관계수의 절댓값이 1에 가까울수록 두 확률변수는 강한 양의 상관관계가 있다.
- ④ 상관계수는 두 확률변수의 선형 관계를 나타내는 측도이다.

문 5. 어떤 전자부품의 불량률은 0.1이다. 전체 생산된 부품에서 임의로 추출된 100개의 부품 가운데 불량품의 개수를 확률변수 X라 할 때, 옳지 않은 것은?

- ① $P(X=10) > P(X=90)$
- ② $P(X \geq 1) = 1$
- ③ X의 기댓값은 10이다.
- ④ X의 표준편차는 3이다.

문 6. 두 확률변수 X, Y에 대한 설명으로 옳지 않은 것은?

- ① X+Y의 기댓값 $E(X+Y) = E(X) + E(Y)$ 이다.
- ② $E(XY) = E(X)E(Y)$ 이면 확률변수 X와 Y는 서로 독립이다.
- ③ 확률변수 X와 Y가 서로 독립이면 두 확률변수의 차 $X-Y$ 의 분산 $Var(X-Y) = Var(X) + Var(Y)$ 이다.
- ④ 확률변수 X와 Y가 서로 독립이면 두 확률변수의 공분산 $Cov(X, Y) = 0$ 이다.

문 7. 모집단에서 크기가 16인 표본을 임의로 추출하여 구한 모평균의 95% 신뢰구간이 (20,25)이다. 다음 설명 중 옳은 것만을 모두 고르면?

- ㄱ. 모평균이 구간 (20,25) 안에 있을 확률은 0.95이다.
- ㄴ. 표본의 95%가 구간 (20,25) 안에 있다.
- ㄷ. 표본평균이 구간 (20,25) 안에 있다.

- ① ㄷ
- ② ㄱ, ㄴ
- ③ ㄱ, ㄷ
- ④ ㄱ, ㄴ, ㄷ

문 8. 평균이 1이고, 분산이 4인 정규모집단에서 추출한 확률표본 X_1, \dots, X_9 의 표본평균이 \bar{X} 이다. 실수 $\alpha (0 < \alpha < \frac{1}{2})$ 에 대해 $P(\bar{X} \leq 1 - 2c) = \alpha$ 를 만족시키는 c의 값은? (단, z_α 는 표준정규분포의 제 $100 \times (1 - \alpha)$ 백분위수이다)

- ① z_α
- ② $z_{1-\alpha}$
- ③ $\frac{1}{3}z_\alpha$
- ④ $\frac{1}{3}z_{1-\alpha}$

문 9. 어느 지역의 2019년 쌀 생산량을 추정하기 위해 이 지역 9개의 농지를 임의로 추출하여 단위면적당 쌀 생산량을 조사하였다니 표본평균 1,150 kg, 표본표준편차 150 kg이다. 이 지역의 2019년 단위면적당 평균 쌀 생산량이 2018년의 단위면적당 평균 쌀 생산량 1,100 kg보다 크다는 주장을 검정하고자 한다. 귀무가설이 참일 때, t 검정통계량의 값과 자유도 k를 옳게 짝 지은 것은? (단, 단위면적당 쌀 생산량은 정규분포를 따른다)

	t 검정통계량의 값	k
①	1	8
②	1	9
③	$\frac{1}{3}$	8
④	$\frac{1}{3}$	9

문 10. 다음 중 서로 독립인 두 확률변수 X_1, X_2 에 대해 옳은 것만을 모두 고르면? (단, \sim 는 확률변수가 해당 확률분포를 따르는 것을 나타낸다)

- ㄱ. $X_1 \sim P(m_1), X_2 \sim P(m_2)$ 이면 두 확률변수의 합 $X_1 + X_2 \sim P(m_1 + m_2)$ 이다. 단, $P(m)$ 은 평균이 m 인 포아송분포
- ㄴ. $X_1 \sim \chi^2(m_1), X_2 \sim \chi^2(m_2)$ 이면 두 확률변수의 합 $X_1 + X_2 \sim \chi^2(m_1 + m_2)$ 이다. 단, $\chi^2(m)$ 은 자유도가 m 인 카이제곱분포
- ㄷ. $X_1 \sim B(m_1, p), X_2 \sim B(m_2, p)$ 이면 두 확률변수의 합 $X_1 + X_2 \sim B(m_1 + m_2, p)$ 이다. 단, $B(m, p)$ 는 베르누이 시행횟수가 m 이고 성공률이 p 인 이항분포
- ㄹ. $X_1 \sim N(\mu_1, \sigma^2), X_2 \sim N(\mu_2, \sigma^2)$ 이면 두 확률변수의 차 $X_1 - X_2 \sim N(\mu_1 - \mu_2, 2\sigma^2)$ 이다. 단, $N(\mu, \sigma^2)$ 은 평균이 μ , 분산이 σ^2 인 정규분포

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

문 11. 다음 귀무가설(H_0)과 대립가설(H_1)에 대한 검정법의 제2종 오류에 해당하는 것은?

$$H_0 : \mu = 10 \text{ 대 } H_1 : \mu \neq 10$$

- ① $\mu = 10$ 일 때 귀무가설을 기각하는 것
- ② $\mu = 15$ 일 때 귀무가설을 기각하는 것
- ③ $\mu = 10$ 일 때 귀무가설을 기각하지 않은 것
- ④ $\mu = 15$ 일 때 귀무가설을 기각하지 않은 것

문 12. 어떤 공무원이 종합민원실에 방문하는 민원인의 수가 요일별로 고르게 분포되어 있다고 생각한다. 이 가설을 확인하고자 4주간 요일별로 방문하는 민원인의 수를 정리하여 다음 표를 작성하였다.

요일	월	화	수	목	금	계
민원인의 수	120	80	80	100	120	500

이 가설의 검정방법에 대한 설명으로 옳지 않은 것은? (단, $\chi^2_k(\alpha)$ 는 자유도가 k 인 카이제곱분포의 제 $100 \times (1 - \alpha)$ 백분위수를 나타낼 때 $\chi^2_{0.05}(4) = 9.49, \chi^2_{0.05}(5) = 11.07$ 이다)

- ① 귀무가설이 참일 때 민원인의 각 요일별 방문 확률은 0.2이다.
- ② 귀무가설이 참일 때 월요일에 방문하는 민원인의 수의 기대도수는 100이다.
- ③ 이 가설을 검정하기 위한 카이제곱통계량의 값은 20이다.
- ④ 유의수준 5%에서 요일별 방문 확률은 같다고 할 수 없다.

문 13. 다음 분산분석에 대한 설명으로 옳은 것만을 모두 고르면?

- ㄱ. 3개 이상 모집단의 평균비교를 위한 통계 기법이다.
- ㄴ. 각 비교 대상 모집단은 서로 다른 분산을 가져야 한다.
- ㄷ. 일원배치법은 각 처리수준에서 반복수가 일정해야 한다.
- ㄹ. 반복이 있는 이원배치법에서는 두 요인의 교호작용을 검정할 수 있다.

- ① ㄱ, ㄷ
- ② ㄱ, ㄹ
- ③ ㄴ, ㄷ
- ④ ㄴ, ㄹ

문 14. 키가 큰 사람의 연봉이 높다는 주장이 있다. 이를 검정하기 위해 MBA 졸업생 250명의 키(X)와 연봉(Y)을 조사하여 단순선형 회귀분석한 결과의 일부가 다음 표와 같다. 키에 대한 회귀계수의 최소제곱추정량의 값 ㉠은? (단, 키에 대한 회귀계수는 양수이다)

<분산분석표>

	자유도	제곱합	평균제곱	F값	p-값
회귀	1	960	960	16	0.00
잔차	248				
계	249				

	계수	표준 오차	t값	p-값
Y 절편	17.93	11.48	1.56	0.12
키	㉠	0.20		0.00

- ① 0.8
- ② 1.0
- ③ 2.0
- ④ 4.0

문 15. 두 인자 A와 B가 어떤 특성값에 영향을 주는가에 대해 인자의 수준 조합에서 동일하게 반복실험을 행하여 분산분석을 한 결과의 일부가 다음 표와 같다. 이에 대한 설명으로 옳지 않은 것은?

<분산분석표>

요인	제곱합	자유도	평균제곱	p-값
인자 A	9	3	3	
인자 B	2			
교호작용 A×B	900	6	150	㉠
오차	20			
계	931	23		

- ① 인자 A의 수준 수는 4이고, 인자 B의 수준 수는 3이다.
- ② 인자 A, B의 수준 조합에서 반드시 두 번의 반복실험을 해야 한다.
- ③ 유의수준 5%에서 ㉠이 0.05보다 작으면 인자 A와 B의 각각에 대한 유의성 검정은 의미가 없다.
- ④ 유의수준 5%에서 ㉠이 0.05보다 작으면 교호작용을 오차항에 풀링(pooling)하여 분산분석을 다시 실시한다.

문 16. 흡연과 심근경색의 연관성을 평가하기 위하여 특정 병원을 방문하여 심근경색으로 진단된 환자 중에서 300명, 심근경색이 아닌 환자 중에서 600명을 각각 임의추출하여 흡연 여부를 조사한 분할표가 다음과 같다.

	심근경색 여부	예	아니오	계
흡연 여부	예	150	150	300
	아니오	150	450	600
	계	300	600	900

- 이 분할표에 대한 설명으로 옳은 것은?
- 동질성 검정의 카이제곱통계량의 자유도는 2이다.
 - 동질성 검정의 카이제곱통계량의 값은 $\chi^2 = \frac{(150-100)^2}{150} + \frac{(150-200)^2}{150} + \frac{(150-200)^2}{150} + \frac{(450-400)^2}{450}$ 이다.
 - 분할표로부터 심근경색이 발병한 환자에서의 흡연율을 추정할 수 있다.
 - 분할표로부터 흡연한 사람에서의 심근경색 발생률을 추정할 수 있다.

문 17. 어떤 질병의 새로운 치료약을 개발하였다. 다음 표는 복용량에 따라 이 질병 증상의 경감기간을 알아보기 위해 10명의 환자에 대하여 실험한 자료의 일부이다.

번호	1	2	3	...	8	9	10
복용량(x)	3	3	4	...	8	8	9
경감기간(y)	9	5	12	...	18	24	22

복용량과 경감기간의 관계를 단순선형회귀모형 $y_i = \alpha + \beta x_i + \epsilon_i$ ($i = 1, 2, \dots, 10$)에 적용하였다. 이때 ϵ_i 는 서로 독립이며 평균이 0, 분산이 σ^2 인 정규분포를 따른다. $\sum_{i=1}^{10} (x_i - \bar{x})^2 = 90$,

$\sum_{i=1}^{10} (y_i - \bar{y})^2 = 1,000$, $\sum_{i=1}^{10} (x_i - \bar{x})(y_i - \bar{y}) = 270$ 일 때, 옳지 않은 것은? (단, \bar{x} , \bar{y} 는 각각 x , y 의 표본평균이다)

- 회귀계수 β 의 최소제곱추정량의 값은 3이다.
- 복용량과 경감기간의 상관계수는 0.9이다.
- 결정계수는 0.81이다.
- σ^2 의 불편추정량의 값은 $\frac{190}{9}$ 이다.

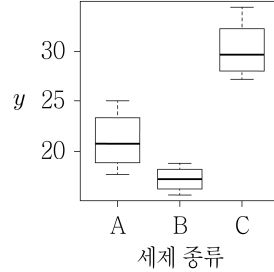
문 18. 다음 표에서 x' 와 y' 는 각각 설명변수(x)와 반응변수(y)를 표준화한 변수 $x' = \frac{x - \bar{x}}{s_x}$, $y' = \frac{y - \bar{y}}{s_y}$ 의 자료이다.

번호	1	2	3	4	5	6	7	8	9	10
x'	-1.7	0.0	0.4	0.0	0.0	1.8	-0.9	0.9	-0.9	0.4
y'	-1.4	0.1	0.5	-0.6	-0.3	1.9	-0.1	1.0	-1.3	0.2

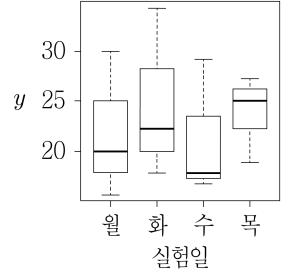
위의 표준화된 변수 (x', y')에 대해 단순선형회귀분석을 최소제곱법으로 수행할 때, 회귀계수의 추정값에 대한 설명으로 옳은 것은? (단, \bar{x} , \bar{y} 는 각각 x , y 의 표본평균이고, s_x , s_y 는 각각 x , y 의 표본표준편차이고, $s_x \neq s_y$ 이다)

- 원 변수 (x, y)에 대한 회귀계수의 추정값과 동일하다.
- 원 변수 (x, y)에 대한 회귀계수의 추정값의 역수와 동일하다.
- 원 변수 (x, y)의 상관계수와 동일하다.
- 원 변수 (x, y)의 결정계수와 동일하다.

문 19. 3가지 종류의 유기농 세제(A, B, C)에 대해 박테리아의 성장을 지연시키는 효과를 비교하고자 한다. 실험실에서 하루에 3번의 실험만 가능하므로 실험일 월, 화, 수, 목을 블록으로 하여 확률화완전블록설계를 실시하여 다음과 같은 결과를 얻었다. 분산분석을 이용하여 세제 종류별 박테리아 성장 지연 효과의 차이를 검정하고자 한다. 분석결과에 대한 설명으로 옳은 것만을 모두 고르면?



<그림 1> 세제 종류별 상자그림



<그림 2> 실험일별 상자그림

<분산분석표>

요인	자유도	제곱합	평균제곱	F값	p-값
세제(처리)		360	180	36	0.00
실험일(블록)		30	10	2	0.22
오차		30	5		
계					

- 유의수준 5%에서 세제 종류별 박테리아의 성장을 지연시키는 효과의 차이는 유의하다.
- 유의수준 5%에서 실험일간 특성값(y)의 산포가 매우 작다고 할 수 있다.
- 귀무가설이 참일 때 세제에 대한 검정통계량의 확률분포는 분자의 자유도는 3, 분모의 자유도는 6인 F 분포이다.
- 세제 종류별 상자그림에서 범위가 가장 큰 세제는 B이다.

- 가, 나
- 가, 다
- 가, 나, 다
- 나, 다, 라

문 20. 어떤 직육면체 주사위의 1의 눈이 있는 면이 작아서 1의 눈이 나올 확률 p 가 $\frac{1}{6}$ 보다 작을 것으로 예상된다. 이에 대한 가설

$H_0: p = \frac{1}{6}$ 대 $H_1: p < \frac{1}{6}$ 을 검정하기 위해 이 주사위를 10번 던져 1의 눈이 나온 횟수를 확률변수 Y 라 할 때, 옳은 것은?

- $\frac{Y}{10}$ 는 p 의 불편추정량(unbiased estimator)이 아니다.
- Y 의 관측값이 클수록 귀무가설을 기각한다.
- Y 의 확률분포는 이항분포 $B\left(10, \frac{1}{6}\right)$ 을 따른다.
- Y 의 관측값이 2일 때 유의확률(p -값)은 $P(X \leq 2)$ 이다. 단, 확률변수 X 는 이항분포 $B\left(10, \frac{1}{6}\right)$ 을 따른다.