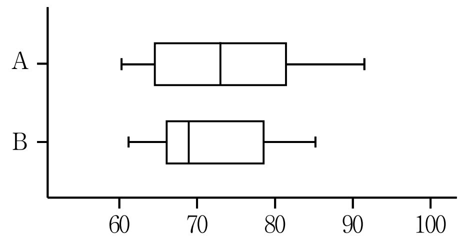
통 계 학

문 1. 다음은 어느 대학교의 학과 A와 B에 대한 통계학 성적의 상자그림(box plot)이다. 이에 대한 설명으로 옳지 않은 것은?



① 성적이 가장 좋은 학생은 학과 A에 속해 있다.
② 학과 A의 사분위범위(interquartile range)는 학과 B의 사분위범위보다 크다.
③ 학과 A가 학과 B보다 70점 이하인 학생의 비율이 높다.
④ 학과 A의 제2사분위수는 학과 B의 제2사분위수보다 크다.

문 2. 다음은 표본의 크기가 17인 어느 자료에 대하여 단순선형회귀모형 $Y_i=\beta_0+\beta_1 X_i+\epsilon_i,\ i=1,2,\cdots,17$을 적용하여 얻은 분산분석표의 일부이다. 이에 대한 설명으로 옳지 않은 것은?

요인	제곱합	자유도	평균제곱	$F-$값	$p-$값
회귀	20			(가)	0.0064
오차(잔차)			(나)		
합계	50				

① Y의 총변동 중 40%는 X로 설명할 수 있다.
② (가)는 $\frac{32}{3}$이다.
③ (나)는 2이다.
④ 유의수준(significance level) 5%에서 X가 Y에 영향을 준다고 할 수 있다.

문 3. 이산확률변수 X의 확률분포(probability distribution)는 다음과 같다. $E(X)=1.6$일 때, 상수 c_1과 c_2의 값은?

x	-1	0	1	2	3	4	5	계
$P(X=x)$	0.05	c_1	0.30	c_2	0.10	0.10	0.05	1

	c_1	c_2
①	0.25	0.15
②	0.2	0.2
③	0.3	0.1
④	-0.2	0.6

문 4. 어떤 회사에서 생산한 건전지의 수명은 평균이 140시간, 표준편차가 20시간인 정규분포를 따른다고 한다. 가설 $H_0:\mu\le 140$ 대 $H_1:\mu>140$을 검정하기 위해 이 회사에서 100개의 건전지를 임의추출하여 조사한 결과 수명의 표본평균이 150시간이었다. 이에 대한 설명으로 옳지 않은 것은? (단, μ는 모평균이고, $z_{0.025}=1.96,\ z_{0.05}=1.645$이다)

① 검정통계량의 값은 5이다.
② 위 가설의 검정은 단측검정이다.
③ 위 가설에 대한 유의확률($p-$값)은 0.05보다 작다.
④ 유의수준 5%에서 귀무가설을 기각할 수 없다.

문 5. A회사 휴대전화기의 결함률은 0.01, B회사는 0.02, C회사는 0.03, D회사는 0.02라고 한다. 각 회사별 휴대전화기의 시장 점유율은 A회사는 0.4, B회사는 0.3, C회사는 0.2, D회사는 0.1이다. 어떤 휴대전화기가 결함을 보였을 때, 이 휴대전화기가 B회사 또는 C회사의 제품일 확률은?

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

문 6. 다음의 두 자료 A, B에 대한 각각의 표본상관계수는?

자료 A	(1, 0.1), (2, 0.2), (3, 0.3), (4, 0.4), (5, 0.5)
자료 B	(1, 0.01), (2, 0.02), (3, 0.03), (4, 0.04), (5, 0.05)

	A	B
①	0.1	0.01
②	0.1	1
③	1	0.1
④	1	1

문 7. 체비셰프 부등식(Chebyshev's inequality)은 "확률변수 X의 평균이 μ, 분산이 $\sigma^2<\infty$일 때, 0보다 큰 모든 k에 대해 $P(|X-\mu|\ge k\sigma)\le\frac{1}{k^2}$"이다. $E(X)=17$이고 $E(X^2)=298$일 때, 체비셰프 부등식을 이용한 $P(11<X<23)$의 하한값은?

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

문 8. 다음은 도시 A와 도시 B에서 지하철 이용 비율에 차이가 있는지 알아보기 위해 두 지역에서 각각 5000명, 3000명을 임의추출하여 조사한 결과이다. 도시 A의 지하철 이용 비율이 도시 B보다 높은지 검정하기 위한 Z 검정통계량의 값과 유의수준 5%에서의 임계치(critical value)는?

도시	출근자 표본 수	지하철 이용 출근자 수
A	5000	4000
B	3000	2000

	Z 검정통계량의 값	임계치
①	$\dfrac{\frac{4}{5}-\frac{2}{3}}{\sqrt{\frac{3}{4}(1-\frac{3}{4})\left(\frac{1}{5000}+\frac{1}{3000}\right)}}$	$z_{0.05}=1.645$
②	$\dfrac{\frac{4}{5}-\frac{2}{3}}{\sqrt{\frac{4}{5}(1-\frac{4}{5})\left(\frac{1}{5000}+\frac{1}{3000}\right)}}$	$z_{0.05}=1.645$
③	$\dfrac{\frac{4}{5}-\frac{2}{3}}{\sqrt{\frac{3}{4}(1-\frac{3}{4})\left(\frac{1}{5000}+\frac{1}{3000}\right)}}$	$z_{0.025}=1.96$
④	$\dfrac{\frac{4}{5}-\frac{2}{3}}{\sqrt{\frac{4}{5}(1-\frac{4}{5})\left(\frac{1}{5000}+\frac{1}{3000}\right)}}$	$z_{0.025}=1.96$

문 9. 다음과 같은 확률분포를 따르는 확률변수 X에 대하여 3개의 임의표본(random sample)으로 2, 2, 3을 얻었을 때, 모수 p에 대한 최대우도추정치(maximum likelihood estimate)는? (단, $0<p<1$이다)

x	1	2	3	계
$P(X=x)$	p^2	$2p(1-p)$	$(1-p)^2$	1

① $\frac{1}{4}$

② $\frac{1}{3}$

③ $\frac{1}{2}$

④ $\frac{2}{3}$

문 10. 어느 학급의 영어 성적이 정규분포를 따른다고 할 때, 크기가 9인 임의표본에서 계산된 모평균의 95% 신뢰구간 길이는 2였다. 신뢰구간의 길이를 0.5 이하가 되도록 하기 위한 최소 표본크기는? (단, 모분산은 알려져 있다)

① 36
② 64
③ 100
④ 144

문 11. 어느 대학교에서 재학생의 의식구조를 조사하기 위해 전체 학과 중 10개 학과를 임의추출한 후, 추출된 학과의 모든 학생들을 표본으로 추출하였다. 여기서 사용한 표본추출방법은? (단, 학과 간의 의식구조는 동질적이고 학과 내에서는 이질적이다)

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

문 12. 어느 실험결과에 대해 일원배치 분산분석법(one-way analysis of variance)을 적용하여 얻은 분산분석표의 일부가 다음과 같을 때, 옳지 않은 것은?

요인	제곱합	자유도	평균제곱	$F-$값
처리	8	(나)	4	(라)
오차(잔차)	(가)	6	(다)	
합계	26			

① (가)는 18이다.
② (나)는 2이다.
③ (다)는 3이다.
④ (라)는 2이다.

문 13. 중심극한정리(central limit theorem)에 대한 설명으로 옳은 것은?

① 이항분포에서 표본크기 n이 크고 성공확률 p가 작으면 근사적으로 모수가 $\lambda=np$인 포아송분포를 따른다.
② 임의의 모집단에서 표본크기 n이 충분히 커지면 표본평균 $\overline{X}=\frac{1}{n}\sum_{i=1}^{n}X_i$는 근사적으로 정규분포를 따른다.
③ 확률실험의 횟수가 증가하면 어떤 사건이 관측되는 비율은 그 사건의 확률로 수렴한다.
④ 표본평균 $\overline{X}=\frac{1}{n}\sum_{i=1}^{n}X_i$와 표본분산 $S^2=\frac{1}{n-1}\sum_{i=1}^{n}(X_i-\overline{X})^2$은 독립이다.

문 14. 두 이산확률변수 X와 Y의 결합확률분포(joint probability distribution)가 다음과 같을 때, 이에 대한 설명으로 옳은 것만을 모두 고른 것은?

X \ Y	0	1	2
-1	0	$\frac{1}{4}$	0
0	$\frac{1}{4}$	0	$\frac{1}{4}$
1	0	$\frac{1}{4}$	0

ㄱ. $E(X)$는 0이다.
ㄴ. $E(X|Y=1)$은 0이다.
ㄷ. X와 Y는 독립이다.

① ㄱ
② ㄱ, ㄴ
③ ㄴ, ㄷ
④ ㄱ, ㄴ, ㄷ

문 15. 피어슨(Pearson)의 표본상관계수(r)에 대한 설명으로 옳은 것만을 모두 고른 것은?

ㄱ. r의 범위는 $-1 \le r \le 1$이다.
ㄴ. r은 두 변수 사이에 존재하는 직선관계 또는 곡선관계의 정도를 측정한다.
ㄷ. r이 0이면 두 변수는 항상 독립이다.

① ㄱ
② ㄱ, ㄴ
③ ㄱ, ㄷ
④ ㄴ, ㄷ

문 16. 다중선형회귀(multiple linear regression)에 대한 설명으로 옳지 않은 것은?

① 결정계수(coefficient of determination)는 총변동 중에서 회귀변동이 차지하는 비율이다.
② 결정계수의 값은 독립변수가 추가될수록 커진다.
③ 다중공선성(multicollinearity)은 독립변수와 종속변수 간의 상관관계가 높은 경우에 발생한다.
④ 다중공선성 유무는 분산확대인자(variance inflation factor)로 검토할 수 있다.

문 17. 새로운 정책에 대해 산업별로 찬반에 차이가 있는지 알아보기 위하여 중공업분야 전문가 130명, 농업분야 전문가 130명을 임의추출하여 조사한 결과가 다음과 같을 때, 이에 대한 설명으로 옳은 것은? (단, 중공업에서의 찬성률은 P_1, 농업에서의 찬성률은 P_2, 비율차는 $P_1 - P_2$, 상대비율은 $\dfrac{P_1}{P_2}$, 오즈비는 $\dfrac{P_1/(1-P_1)}{P_2/(1-P_2)}$이다)

산업 분야 \ 정책 찬반	찬성	반대	계
중공업	100	30	130
농업	70	60	130
계	170	90	260

① 표본오즈비는 $\dfrac{100 \times 70}{30 \times 60} = \dfrac{35}{9}$이다.
② 비율차가 0에 가까울수록 두 분야의 찬성률 차이는 커질 것이다.
③ 상대비율이 1에 가까울수록 두 분야의 찬성률 차이는 작아질 것이다.
④ 오즈비의 값이 0.5이면 두 분야의 찬성률은 동일하다고 할 수 있다.

문 18. 다음은 어느 지역에서 600명을 임의추출하여 흡연 여부와 치주질환 유무를 조사한 결과이다. 흡연 여부와 치주질환 유무에 대한 독립성 검정을 하고자 할 때, 이에 대한 설명으로 옳은 것만을 모두 고른 것은?

흡연 여부 \ 치주질환 유무	있음	없음	계
과거흡연	50	100	150
현재흡연	50	100	150
비흡연	100	200	300
계	200	400	600

ㄱ. 귀무가설은 “흡연 여부와 치주질환 유무는 서로 관계가 있다.”이다.
ㄴ. 카이제곱 검정통계량의 값은 0이다.
ㄷ. 검정통계량 값이 임계치보다 작으면 귀무가설을 기각한다.
ㄹ. 귀무가설 하에서 카이제곱 검정통계량의 자유도는 2이다.

① ㄱ, ㄹ
② ㄴ, ㄷ
③ ㄱ, ㄷ
④ ㄴ, ㄹ

문 19. 관측쌍 $(X_1, Y_1), (X_2, Y_2), \cdots, (X_n, Y_n)$의 대응(쌍체)비교 t−검정(paired t-test)에 대한 설명 중 옳지 않은 것은?

① 각 관측쌍들 간은 독립이고, 쌍 내의 자료들은 독립이 아닌 경우에 사용된다.
② 쌍으로 관측된 자료의 차에 대해 일표본 t−검정을 적용하는 방법이다.
③ 귀무가설 하에서 검정통계량의 자유도는 $2(n-1)$이다.
④ n이 충분히 큰 경우에는 정규분포로 근사하여 검정할 수 있다.

문 20. X_1과 X_2가 다음과 같은 확률밀도함수(probability density function)로부터 얻어진 임의표본일 때, 확률 $P(\min(X_1, X_2) < \frac{1}{2})$의 값은? (단, $\min(X_1, X_2)$는 X_1과 X_2 중에서 더 작은 확률변수이다)

$$f(x) = 4x^3,\ \ 0 < x < 1$$

① $\dfrac{3}{8}$
② $\dfrac{1}{4}$
③ $\dfrac{21}{128}$
④ $\dfrac{31}{256}$