통계학개론

문 1. 다음은 자료 $x_1, x_2, \cdots, x_n$에 대한 표본평균($\overline{x}$)의 성질이다. 이에 대한 설명으로 옳은 것만을 모두 고르면?

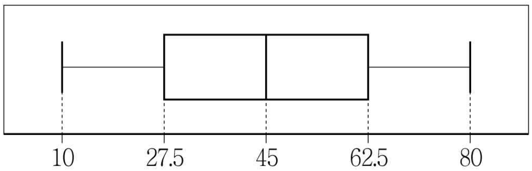
> ㄱ. 5개의 자료 1, 3, 5, 7, 9의 표본평균은 5이다.
> ㄴ. 표본평균보다 값이 큰 자료 수와 작은 자료 수는 같다.
> ㄷ. $\sum_{i=1}^{n}\left(x_i - \overline{x}\right)$의 값은 항상 0이다.

① ㄱ, ㄴ ② ㄱ, ㄷ
③ ㄴ, ㄷ ④ ㄱ, ㄴ, ㄷ

문 2. $P(A)=\frac{1}{3},\ P(B)=\frac{1}{4},\ P(B|A)+P(A|B)=\frac{7}{12}$일 때, $P(A\cap B)$의 값은?

① $\frac{1}{12}$ ② $\frac{1}{6}$
③ $\frac{1}{4}$ ④ $\frac{1}{3}$

문 3. 다음 상자그림(box plot)에 대한 설명으로 옳지 않은 것은?



① 사분위수 범위는 35이다.
② 45이상의 값을 갖는 자료는 전체 자료의 35%이다.
③ 범위(range)는 70이다.
④ 최솟값은 10이다.

문 4. $X_1, X_2, \cdots, X_n$은 평균이 μ, 분산이 σ^2인 확률표본(random sample)이라고 하자. 표본평균 $\overline{X}$에 대한 설명 중 옳은 것만을 모두 고르면?

> ㄱ. $\overline{X}$의 분산은 X_1의 분산보다 크다.
> ㄴ. $\overline{X}$의 기댓값은 X_1의 기댓값과 같다.
> ㄷ. $\overline{X}$의 분산은 n이 커질수록 작아진다.

① ㄱ, ㄴ ② ㄱ, ㄷ
③ ㄴ, ㄷ ④ ㄱ, ㄴ, ㄷ

문 5. 세 확률변수 X_1, X_2, X_3은 서로 독립이고, 각각의 확률분포는 다음과 같다. 변환된 확률변수 $W=X_1-2X_2+X_3$에 대해 $P(W>a)=0.05$를 만족하는 실수 a의 값은? (단, z_α는 표준정규분포의 제$100\times(1-\alpha)$ 백분위수이다)

> $X_1 \sim N(2, 1^2),\ X_2 \sim N(1, 1^2),\quad X_3 \sim N(2, 2^2)$

① $2+3\times z_{0.05}$
② $2+3\times z_{0.95}$
③ $2+4\times z_{0.05}$
④ $2+4\times z_{0.95}$

문 6. 어느 보험회사에서 도시 근로자의 평균 나이(μ)를 추정하기 위하여 64명을 임의로 추출하여 조사하였다. 64명 도시 근로자의 평균 나이가 36.38이고 표준편차가 11.07일 때, 모평균 μ에 대한 95% 신뢰구간은? (단, 표준정규분포를 따르는 확률변수 Z에 대하여 $P(Z\geq 1.96)=0.025$, $P(Z\geq 1.645)=0.05$이다)

① $\left(36.38-1.645\times\frac{11.07}{64},\quad 36.38+1.645\times\frac{11.07}{64}\right)$
② $\left(36.38-1.645\times\frac{11.07}{8},\quad 36.38+1.645\times\frac{11.07}{8}\right)$
③ $\left(36.38-1.96\times\frac{11.07}{64},\quad 36.38+1.96\times\frac{11.07}{64}\right)$
④ $\left(36.38-1.96\times\frac{11.07}{8},\quad 36.38+1.96\times\frac{11.07}{8}\right)$

문 7. 다음은 정부의 미세먼지 관련 정책에 대한 남녀 간 지지도의 차이를 알아보기 위해서 남녀 각각 100명씩을 조사하여 얻은 결과표이다. 남성의 지지율(p_1)보다 여성의 지지율(p_2)이 더 큰지 유의수준 5%에서 검정할 때, Z 검정통계량의 값(Z_0)과 기각역을 옳게 짝 지은 것은? (단, z_α는 표준정규분포의 제$100\times(1-\alpha)$ 백분위수이다)

구분		지지 여부		합계
		지지함	지지하지 않음	
성별	남성	40	60	100
	여성	60	40	100
합계		100	100	200

<u>Z 검정통계량</u> <u>기각역</u>

① $Z_0=\dfrac{0.4-0.6}{\sqrt{\dfrac{0.5\times0.5}{100}+\dfrac{0.5\times0.5}{100}}}$ $Z_0\geq z_{0.05}$

② $Z_0=\dfrac{0.4-0.6}{\sqrt{\dfrac{0.4\times0.6}{100}+\dfrac{0.6\times0.4}{100}}}$ $Z_0\geq z_{0.05}$

③ $Z_0=\dfrac{0.4-0.6}{\sqrt{\dfrac{0.5\times0.5}{100}+\dfrac{0.5\times0.5}{100}}}$ $Z_0\leq -z_{0.05}$

④ $Z_0=\dfrac{0.4-0.6}{\sqrt{\dfrac{0.4\times0.6}{100}+\dfrac{0.6\times0.4}{100}}}$ $Z_0\leq -z_{0.05}$

문 8. 두 확률변수 X, Y의 상관계수에 대한 설명으로 옳은 것만을 모두 고르면?

> ㄱ. X와 Y의 상관계수가 0이면 X와 Y가 서로 독립이다.
> ㄴ. X와 Y가 서로 독립이면 상관계수가 0이다.
> ㄷ. $P\left(Y=\frac{X}{2}+1\right)=1$이면 X와 Y의 상관계수는 $\frac{1}{2}$이다.

① ㄴ
② ㄱ, ㄴ
③ ㄱ, ㄷ
④ ㄴ, ㄷ

문 9. 다음은 카이제곱통계량을 이용하여 두 변수가 서로 독립인지 알아보기 위한 관측도수의 2×2 분할표이다. 카이제곱(χ^2) 검정에 대한 설명으로 옳지 않은 것은? (단, 귀무가설이 참일 때 각 셀의 기대도수는 5 이상이고, 카이제곱통계량의 값은 k이다)

구분		변수2		합계
		범주1	범주2	
변수1	범주1	O_{11}	O_{12}	n_{1+}
	범주2	O_{21}	O_{22}	n_{2+}
합계		n_{+1}	n_{+2}	n

① 전체 관측도수의 합과 전체 기대도수의 합은 같다.
② X가 자유도 1인 카이제곱분포를 따를 때, 유의확률은 $P(X\le k)$와 같다.
③ 관측도수가 O_{11}인 셀의 기대도수와 O_{12}인 셀의 기대도수의 합은 n_{1+}와 같다.
④ 관측도수가 O_{11}인 셀의 기대도수는 $\dfrac{(n_{1+})\times(n_{+1})}{n}$과 같다.

문 10. $F-$분포에 대한 설명으로 옳은 것만을 모두 고르면? (단, $F_\alpha(k_1, k_2)$는 분자의 자유도가 k_1이고 분모의 자유도가 k_2인 $F-$분포의 제$100\times(1-\alpha)$ 백분위수이다)

ㄱ. 자유도 k_1, k_2에 대해 항상 $F_\alpha(k_1,k_2)\times F_{1-\alpha}(k_2,k_1)=1$이다.
ㄴ. T가 자유도 k인 $t-$분포를 따를 때, 확률변수 $\dfrac{1}{T^2}$은 분자의 자유도가 k이고 분모의 자유도가 1인 $F-$분포를 따른다.
ㄷ. 서로 독립인 두 확률변수 Z_1과 Z_2가 표준정규분포를 따를 때, 확률변수 $\left(\dfrac{Z_1}{Z_2}\right)^2$은 분자의 자유도가 1이고 분모의 자유도가 1인 $F-$분포를 따른다.

① ㄱ, ㄴ
② ㄱ, ㄷ
③ ㄴ, ㄷ
④ ㄱ, ㄴ, ㄷ

문 11. 중심극한정리에 대한 설명으로 ㉠, ㉡에 들어갈 말을 옳게 짝 지은 것은? (단, 모집단의 평균이 μ이고, 분산 σ^2은 존재한다)

표본크기가 충분히 클 때, 임의의 분포에서 추출한 확률표본의 (㉠)은 근사적으로 (㉡)를 따른다.

	㉠	㉡
①	표본평균	카이제곱분포
②	표본평균	균등분포
③	표준화 표본평균	지수분포
④	표준화 표본평균	표준정규분포

문 12. 다음은 세 가지 속독법(A, B, C)에 따라 책 읽는 시간에 차이가 있는지 알아보기 위해 일원배치분산분석법을 적용하여 얻은 분산분석표이다. 각 속독법에 5 명씩 15 명을 임의로 배치하여 책을 읽게 한 후, 책 읽는 시간을 측정하였다. 이에 대한 설명으로 옳지 않은 것은?

구분	제곱합	자유도	평균제곱	$F-$값	$p-$값
처리	2156	(㉠)		16.84	0.0003
오차	768	(㉡)	(㉢)		
합계	2924	14			

① ㉠의 값은 3이다.
② ㉡의 값은 12이다.
③ ㉢의 값은 64이다.
④ 유의수준 1%에서 검정할 때, 세 가지 속독법에 따라 책 읽는 시간에 차이가 있다고 할 수 있다.

문 13. 다음은 입학 시 수학 성적(X)과 1학년 때의 통계학 성적(Y)에 대하여 단순선형회귀모형 $Y_i=\alpha+\beta X_i+\epsilon_i,\ i=1,2,\cdots,n$을 적용하여 얻은 결과이다. 이에 대한 설명으로 옳은 것은? (단, $F_\alpha(k_1,k_2)$는 분자의 자유도가 k_1이고 분모의 자유도가 k_2인 $F-$분포의 제$100\times(1-\alpha)$ 백분위수를 나타내고, $F_{0.05}(1,10)=4.96,\ F_{0.05}(1,11)=4.84$이다. 그리고 $t_\alpha(k)$는 자유도가 k인 $t-$분포의 제$100\times(1-\alpha)$ 백분위수를 나타내고, $t_{0.05}(10)=1.812,\ t_{0.025}(10)=2.228,\ t_{0.025}(11)=2.201$이다)

요인	제곱합	자유도	평균제곱	$F-$값
회귀	541.69	1	541.69	29.04
잔차	186.56	10	18.66	

	회귀계수	표준오차	t-값
상수항	30.04	10.14	2.96
X	0.90	0.17	5.34

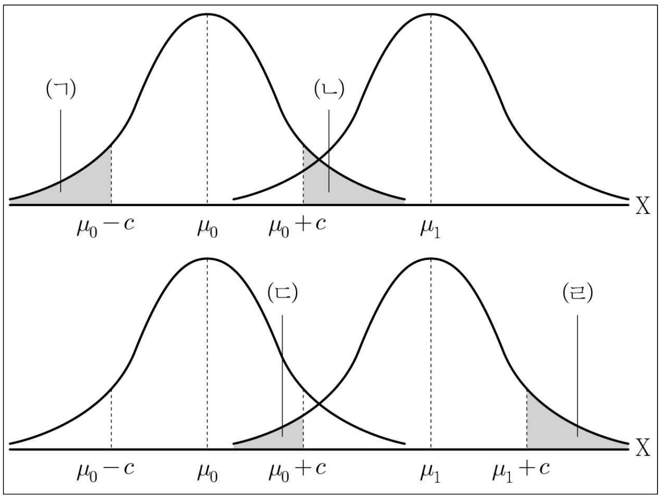
① 추정된 회귀모형의 유의성을 검정할 때, 귀무가설(H_0: 회귀모형은 유의하지 않다)은 유의수준 5%에서 기각된다.
② X와 Y 사이의 모상관계수(ρ)가 0인지 검정할 때, 귀무가설 ($H_0:\rho=0$)은 유의수준 5%에서 기각되지 않는다.
③ 추정된 회귀직선은 $\hat{Y}=10.14+0.17X$이다.
④ 자료의 개수(n)는 11이다.

문 14. 확률변수 X와 Y의 분산과 공분산은 다음과 같다. 확률변수 W와 T를 각각 $W=2X+2,\ T=-Y+1$이라고 할 때, W와 T의 상관계수는?

$$V(X)=25,\ V(Y)=16,\ Cov(X,Y)=-10$$

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

문 15. 확률변수 X는 $N(\mu,\ 1)$를 따를 때, 가설 H_0: $\mu=\mu_0$ 대 H_1: $\mu=\mu_1$에 대한 기각역이 $R=\{x:x\geq\mu_0+c\}$로 주어진 경우, 다음 그림에서 제1종 오류를 범할 확률에 해당하는 영역(A)과 제2종 오류를 범할 확률에 해당하는 영역(B)을 옳게 짝 지은 것은? (단, $\mu_1>\mu_0$이고, $c>0$이다)



	A	B
①	ㄱ	ㄷ
②	ㄱ	ㄹ
③	ㄴ	ㄷ
④	ㄴ	ㄹ

문 16. 다음은 다이어트 종류에 따라 체중 감량 효과에 차이가 있는지 알아보기 위해 분산분석을 시행한 결과표이다. 이 결과에서 알 수 있는 내용으로 옳지 않은 것은? (단, $F_\alpha(k_1,\ k_2)$는 분자의 자유도가 k_1이고, 분모의 자유도가 k_2인 $F-$분포의 제$100\times(1-\alpha)$ 백분위수를 나타내고, $F_{0.05}(3,26)=2.98$, $F_{0.025}(3,26)=3.67$이다)

요인	제곱합	자유도	평균제곱	$F-$값
다이어트	6	3		20
오차	2.6			
합계	8.6	29		

① 유의수준 5%에서 다이어트 종류에 따라 체중 감량 효과에 차이가 있다고 할 수 있다.
② $F-$값과 분자의 자유도 3, 분모의 자유도가 26인 $F-$분포를 이용하여 유의확률($p-$값)을 구할 수 있다.
③ $F-$값은 오차의 평균제곱을 처리의 평균제곱으로 나눈 값이다.
④ 다이어트 종류는 4가지이다.

문 17. 자료의 수가 n인 표본 $(x_i,\ y_i)$ $(i=1,\ 2,\cdots,n)$에 대해 다음 두 회귀모형 M_1과 M_2를 적용하여 분석하고자 한다. 두 모형에 대한 설명으로 옳은 것만을 모두 고르면?

M_1: $Y_i=\alpha+\epsilon_i$,
M_2: $Y_i=\alpha+\beta X_i+\epsilon_i$ $(i=1,\ 2,\cdots,n)$

ㄱ. 모형 M_1에서 $\widehat{Y_i}=\overline{Y}$이다.
ㄴ. 모형 M_2의 결정계수는 0 이상이다.
ㄷ. 모형 M_2의 회귀제곱합은 모형 M_1의 회귀제곱합보다 크거나 같다.

① ㄱ, ㄴ　② ㄱ, ㄷ
③ ㄴ, ㄷ　④ ㄱ, ㄴ, ㄷ

문 18. 어떤 자판기에서 판매되는 음료수 용량은 모평균이 $\mu\,(mL)$이고, 모표준편차가 $5mL$인 확률분포를 따른다고 한다. 이 자동판매기에서 임의로 추출한 100개 음료수의 표본평균이 $150mL$일 때, 가설 $H_0:\mu=\mu_0$ 대 $H_1:\mu\neq\mu_0$에 대한 유의수준 α에서 귀무가설을 기각하지 못하는 μ_0의 범위는? (단, z_α는 표준정규분포의 제$100\times(1-\alpha)$ 백분위수이다)

① $\left(150-\frac{1}{2}z_\alpha,\quad 150+\frac{1}{2}z_\alpha\right)$
② $\left(150-\frac{1}{2}z_{\alpha/2},\quad 150+\frac{1}{2}z_{\alpha/2}\right)$
③ $\left(150-\frac{1}{4}z_\alpha,\quad 150+\frac{1}{4}z_\alpha\right)$
④ $\left(150-\frac{1}{4}z_{\alpha/2},\quad 150+\frac{1}{4}z_{\alpha/2}\right)$

문 19. 다음은 금연 프로그램에 참석한 120명을 대상으로 직업군에 따라 금연 성공률에 차이가 있는지 조사한 분할표이다. 이에 대한 설명으로 옳은 것은? (단, $O_{ij}(i=1,\ 2,\ 3,\ 4,\ j=1,\ 2)$는 $(i,\ j)$ 셀에서 얻어진 관측도수이고, $E_{ij}(i=1,\ 2,\ 3,\ 4,\ j=1,\ 2)$는 귀무가설이 참일 때 $(i,\ j)$ 셀에서 얻어진 기대도수이다. $\chi^2_\alpha(k)$는 자유도가 k인 카이제곱분포의 제$100\times(1-\alpha)$ 백분위수이고, χ^2은 검정통계량이다)

구분		금연		합계
		성공함(1)	성공하지 못함(2)	
직업군	사무직(1)	15	15	30
	자영업(2)	15	10	25
	교육관련(3)	12	18	30
	노동직(4)	18	17	35
합계		60	60	120

① 유의수준 5%에서 검정할 때, 기각역은 $\chi^2\geq\chi^2_{0.025}(2)$이다.
② 카이제곱 검정통계량은 $\chi^2=\sum_{i=1}^{4}\sum_{j=1}^{2}\frac{(O_{ij}-E_{ij})^2}{O_{ij}}$이다.
③ $E_{11}=15$이다.
④ 각 직업의 성공률을 $p_i(i=1,\ 2,\ 3,\ 4)$라고 할 때, 귀무가설은 $H_0:p_1=p_2=p_3=p_4=\frac{1}{4}$이다.

문 20. 다음 설명 중 옳은 것만을 모두 고르면?

ㄱ. 유의확률($p-$값)이 유의수준보다 작을 때 귀무가설을 기각한다.
ㄴ. 모수 θ에 관한 불편추정량(unbiased estimator)의 기댓값은 θ이다.
ㄷ. 검정에서 제1종 오류의 확률을 줄이면 제2종 오류의 확률도 줄어든다.

① ㄱ　② ㄱ, ㄴ
③ ㄴ, ㄷ　④ ㄱ, ㄴ, ㄷ