통 계 학

문 1. 다음 사례에 해당하는 표본추출방법은?

> 서울특별시의 25개 각 구로부터 20개씩의 가구를 임의로 뽑아 가구주에 대해 환경문제에 관한 면접조사를 실시하기로 했다.

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

문 2. 다음은 기초통계학을 수강한 학생 20명의 성적을 줄기－잎 그림으로 나타낸 것이다. 이에 대한 설명으로 옳지 않은 것은?

5	2
6	34778
7	23466668
8	23446
9	0

① 범위는 38이다.
② 중앙값은 76이다.
③ 최빈값은 76이다.
④ 제1 사분위수는 82.5이다.

문 3. 관측값 $x_1, x_2, \cdots, x_n$의 표본평균을 $\overline{x}$, 표본분산을 $s^2 = \dfrac{1}{n-1}\sum_{i=1}^{n}(x_i - \overline{x})^2$으로 정의할 때, $n=10$인 경우 $\dfrac{10\sum_{i=1}^{10}x_i^2 - \left(\sum_{i=1}^{10}x_i\right)^2}{s^2}$의 값은? (단, $x_1, x_2, \cdots, x_{10}$의 값이 모두 같지는 않다)

① 9
② 10
③ 90
④ 100

문 4. 자유도가 5인 카이제곱분포에 대한 설명으로 옳은 것은?

① 기댓값은 중앙값보다 크다.
② 왜도(skewness)는 음수이다.
③ 기댓값은 4이다.
④ 분산은 5이다.

문 5. 수명이 평균 100일인 지수분포를 따르는 트랜지스터를 100일 동안 사용했을 때, 앞으로 100일 더 사용할 수 있을 확률은? (단, 평균이 μ인 지수분포의 확률밀도함수는 다음과 같다)

$$f(x) = \begin{cases} \dfrac{1}{\mu}e^{-\frac{1}{\mu}x}, & x \geq 0 \\ 0, & \text{그 외} \end{cases} \qquad (\mu > 0)$$

① $e^{-1/2}$
② e^{-1}
③ e^{-2}
④ e^{-3}

문 6. 어느 공장에서 생산하는 볼트의 지름(단위: mm)은 $N(14, 0.3^2)$, 너트의 지름(단위: mm)은 $N(15, 0.4^2)$을 따르고 볼트의 지름과 너트의 지름은 서로 독립이라고 하자. 이 공장에서 볼트와 너트를 임의로 하나씩 선택할 때, 볼트가 너트에 들어가지 않을 확률은? (단, $\Phi(x)$는 표준정규분포의 누적분포함수이다)

① $\Phi(-2)$
② $\Phi(2)$
③ $\Phi(-1)$
④ $\Phi(1)$

문 7. 평균과 분산이 알려지지 않은 정규모집단으로부터 16개의 표본을 임의추출한 결과 표본평균 $\overline{x}=13$, 표본분산 $s^2=\frac{1}{15}\sum_{i=1}^{16}(x_i-\overline{x})^2=9$를 얻었다. 모평균 μ에 대한 가설 $H_0:\mu=15$ 대 $H_1:\mu\neq 15$를 검정할 때, $p-$값은? (단, Z와 T는 각각 표준정규분포와 자유도가 15인 $t-$분포를 따르는 확률변수이다)

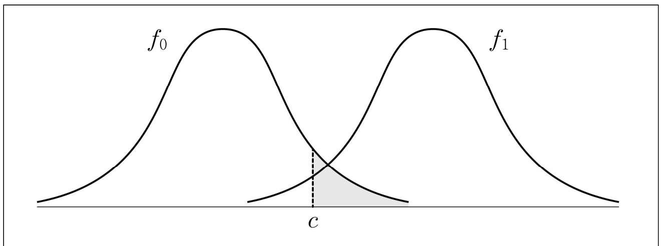
① $2P\left(Z\geq\frac{2\sqrt{15}}{3}\right)$

② $2P\left(Z\geq\frac{8}{3}\right)$

③ $2P\left(T\geq\frac{2\sqrt{15}}{3}\right)$

④ $2P\left(T\geq\frac{8}{3}\right)$

문 8. 다음은 가설 $H_0:\mu=\mu_0$에 대한 유의수준 α인 단측검정에 관한 그림이다. 그림에서 f_0와 f_1으로 표시된 분포는 각각 귀무가설과 대립가설의 특정 모숫값하에서의 검정통계량의 분포를 나타내며, c는 검정의 임곗값(critical value)을 나타낸다. 이에 대한 설명으로 옳은 것만을 모두 고른 것은?



ㄱ. 대립가설은 $H_1:\mu<\mu_0$이다.
ㄴ. 음영 부분의 넓이는 α이다.
ㄷ. f_1이 우측으로 이동할수록 검정력은 커진다.
ㄹ. c를 검정통계량의 값으로 정할 때 음영부분의 넓이가 $p-$값이다.

① ㄱ, ㄹ
② ㄴ, ㄷ
③ ㄴ, ㄹ
④ ㄴ, ㄷ, ㄹ

문 9. 표준정규분포와 자유도가 r인 $t-$분포에 대한 설명으로 옳지 않은 것은? (단, $r>2$)

① 두 분포 모두 평균이 0이고, 평균에 대하여 대칭이다.
② $t-$분포의 분산은 1보다 크다.
③ $t-$분포의 제95 백분위수는 표준정규분포의 제95 백분위수보다 작다.
④ 자유도 r이 무한대로 커지면 $t-$분포는 표준정규분포로 수렴한다.

문 10. $X_1,X_2,\cdots,X_{10}$을 $N(\mu,1)$으로부터의 확률표본이라고 하자. 가설 $H_0:\mu=5$ 대 $H_1:\mu\neq 5$를 유의수준 α에서 검정할 때, $\mu=4$에서의 검정력은? (단, z_α와 $\Phi(x)$는 각각 표준정규분포의 제$100\times(1-\alpha)$ 백분위수와 누적분포함수를 나타낸다)

① $\Phi\left(\sqrt{10}-z_{\alpha/2}\right)$
② $1-\Phi\left(\sqrt{10}-z_{\alpha/2}\right)$
③ $\Phi\left(\sqrt{10}+z_{\alpha/2}\right)+\Phi\left(\sqrt{10}-z_{\alpha/2}\right)$
④ $1-\Phi\left(\sqrt{10}+z_{\alpha/2}\right)+\Phi\left(\sqrt{10}-z_{\alpha/2}\right)$

문 11. 영수는 편의점에서 아메리카노 4개, 카페라떼 3개를 구매하여 상자 하나에 담아 두었다. 그 상자에서 임의로 두 개의 커피를 꺼낼 때, 아메리카노가 한 개 이하일 확률은?

① $\frac{5}{7}$
② $\frac{6}{7}$
③ $\frac{5}{8}$
④ $\frac{3}{4}$

문 12. 당뇨병, 고혈압, 고지혈증에 대한 결합확률분포가 다음과 같다. 고지혈증이 있는 환자 중에 고혈압이 없을 조건부 확률 P(고혈압 무 | 고지혈증 유)는?

	당뇨병 유		당뇨병 무	
	고지혈증 유	고지혈증 무	고지혈증 유	고지혈증 무
고혈압 유	0.10	0.01	0.10	0.02
고혈압 무	0.02	0.05	0.10	0.60

① $\frac{5}{16}$

② $\frac{3}{8}$

③ $\frac{7}{16}$

④ $\frac{11}{16}$

문 13. 두 지역 A, B에서 새로운 복지 제도에 대한 찬성 여부를 조사한 결과표 일부가 다음과 같다. 지역과 찬성 여부가 독립이라고 가정할 때 (가)의 값은?

지역 / 찬성 여부	지역 A	지역 B
찬성	25	35
반대	15	(가)

① 10

② 21

③ 25

④ 30

문 14. 어느 지역에서 남자와 여자 각각 150명을 임의추출하여 특정 도시개발계획에 대한 찬성 여부를 조사한 결과가 다음과 같다.

찬성 여부 / 성별	찬성	반대	합계
남자	80	70	150
여자	70	80	150

다음 가설을 유의수준 α에서 검정하고자 할 때, 검정 방법과 임곗값(critical value)을 옳게 짝지은 것은? (단, z_α와 $\chi^2_\alpha(r)$는 각각 표준정규분포와 자유도가 r인 카이제곱분포의 제$100\times(1-\alpha)$ 백분위수이다)

H_0 : 남자의 찬성률 $=$ 여자의 찬성률
H_1 : 남자의 찬성률 $>$ 여자의 찬성률

① (Z 검정, $z_{\alpha/2}$)

② (Z 검정, z_α)

③ (χ^2 검정, $\chi^2_{\alpha/2}(1)$)

④ (χ^2 검정, $\chi^2_\alpha(2)$)

문 15. 여러 종류의 직물에 대해 인화 시간(단위: 초)을 측정한 자료를 이용하여 직물에 따라 인화 시간의 모평균이 다른지를 검정하기 위한 일원배치 분산분석의 결과가 다음과 같을 때, 이에 대한 설명으로 옳지 않은 것은? (단, $F_\alpha(k_1, k_2)$는 자유도가 (k_1, k_2)인 $F-$분포의 제$100\times(1-\alpha)$ 백분위수를 나타내고, $F_{0.01}(2,35)=5.27$, $F_{0.05}(2,35)=3.27$, $F_{0.1}(2,35)=2.46$이다)

요인	제곱합	자유도	평균제곱	$F-$값
직물	72	2		
오차	315	35		
전체	387	37		

① 이 실험에 사용된 직물의 종류는 3개이다.

② 직물에 대한 $F-$값은 4이다.

③ 유의수준 10 %에서 직물에 따라 인화 시간에 차이가 없다.

④ 이 가설검정에 대한 $p-$값은 0.01보다 크고 0.05보다 작다.

문 16. 다음 상황에 대한 검정 방법으로 옳지 않은 것은? (단, 각 검정에 대해 정규성과 등분산성을 만족하고 검정에 필요한 분산은 알려지지 않은 것으로 한다)

> 새로 개발된 당뇨치료제의 효과를 알아보기 위한 임상시험에서 자발적으로 지원한 20명의 당뇨 환자 중 무작위로 10명을 선택하여 A 그룹으로 배정하고 나머지를 B 그룹으로 배정한 후, A 그룹에는 개발된 치료제를, B 그룹에는 위약(placebo)을 한 달간 복용하게 하였다. 각 그룹에 대한 치료 효과를 비교하기 위해 복용 전의 혈당량과 복용 후의 혈당량을 측정하였다.

① A 그룹과 B 그룹의 복용 전 평균 혈당량의 비교는 독립 $t-$검정으로 실시한다.
② A 그룹 내에서 각 환자의 복용 전과 후의 혈당량 비교는 대응(쌍체) $t-$검정으로 실시한다.
③ 각 환자의 복용 전과 후의 혈당량 차이에 대한 A 그룹과 B 그룹 간의 평균 비교는 독립 $t-$검정으로 실시한다.
④ 복용 후 혈당량에 대한 A 그룹과 B 그룹 간의 평균 비교는 대응(쌍체) $t-$검정으로 실시한다.

문 17. 단순선형회귀모형 $Y_i=\beta_0+\beta_1x_i+\epsilon_i,\ i=1,2,\cdots,n$에서 β_0와 β_1의 최소제곱추정량을 각각 $\hat{\beta}_0$와 $\hat{\beta}_1$이라 하고 잔차 e_i를 $Y_i-\hat{\beta}_0-\hat{\beta}_1x_i$이라 할 때, 이에 대한 설명으로 옳은 것만을 모두 고른 것은? (단, ϵ_i는 $N(0,\sigma^2)$을 따르고 서로 독립이다)

> ㄱ. e_i와 x_i의 곱의 합은 0이다.
> ㄴ. $\hat{\beta}_1$의 부호는 설명변수와 반응변수 간의 표본상관계수의 부호와 같다.
> ㄷ. 결정계수의 부호와 $\hat{\beta}_1$의 부호는 같다.
> ㄹ. 결정계수의 부호는 설명변수와 반응변수 간의 표본 상관계수의 부호와 같다.

① ㄱ, ㄴ
② ㄱ, ㄷ
③ ㄴ, ㄹ
④ ㄷ, ㄹ

문 18. 다음 중 선형회귀모형을 적용할 수 있는 모형만을 모두 고른 것은? (단, ϵ_i는 $N(0,\sigma^2)$을 따르고 서로 독립이다)

> ㄱ. $Y_i=\beta_0+\beta_1x_i+\epsilon_i$
> ㄴ. $Y_i=e^{\beta_0}\times x_i^{\beta_1}\times e^{\epsilon_i}$
> ㄷ. $Y_i=\beta_0+\beta_1x_i^2+\epsilon_i$

① ㄱ
② ㄱ, ㄷ
③ ㄴ, ㄷ
④ ㄱ, ㄴ, ㄷ

문 19. 다중선형회귀모형 $Y_i=\beta_0+\beta_1x_{1i}+\beta_2x_{2i}+\cdots+\beta_kx_{ki}+\epsilon_i$, $i=1,2,\cdots,n$에서 결정계수(R^2)와 수정결정계수(adjusted R^2)에 대한 설명으로 옳은 것은? (단, $k\ge 1$)

① 동일한 설명변수를 사용할 때 수정결정계수는 결정계수보다 작거나 같다.
② 설명변수를 추가할수록 결정계수는 작아진다.
③ 설명변수를 추가할수록 수정결정계수는 커진다.
④ 수정결정계수는 다중공선성의 판단기준이 된다.

문 20. 다중선형회귀모형 $Y_i=\beta_0+\beta_1x_{1i}+\beta_2x_{2i}+\epsilon_i,\ i=1,2,\cdots,27$에 최소제곱추정법을 적용하여 얻은 분산분석표가 다음과 같을 때, 이에 대한 설명으로 옳은 것만을 모두 고른 것은? (단, ϵ_i는 $N(0,\sigma^2)$을 따르고 서로 독립이며, 자유도가 (k_1,k_2)인 $F-$분포의 제$100\times(1-\alpha)$ 백분위수를 $F_\alpha(k_1,k_2)$라고 할 때 $F_{0.01}(2,24)=5.61$이다)

요인	제곱합	자유도	평균제곱	$F-$값
회귀	20	2	10	
잔차	15	24	0.625	
계	35	26		

> ㄱ. 회귀직선의 결정계수는 $\dfrac{20}{35}$이다.
> ㄴ. 유의수준 1 %에서 가설 $H_0:\beta_1=0$ 대 $H_1:\beta_1\ne 0$에 대한 검정에서 귀무가설을 기각한다.
> ㄷ. 모형의 유의성 검정에서 귀무가설 모형은 $Y_i=\epsilon_i$이다.

① ㄱ
② ㄱ, ㄷ
③ ㄴ, ㄷ
④ ㄱ, ㄴ, ㄷ

문 21. 단순선형회귀모형 $Y_i=\beta_0+\beta_1(x_i-\overline{x})+\epsilon_i,\ i=1,2,\cdots,n$에서 회귀계수 β_0와 β_1에 대한 최소제곱추정량 $\hat{\beta}_0$과 $\hat{\beta}_1$에 대한 설명으로 옳지 않은 것은? (단, ϵ_i는 $N(0,\sigma^2)$을 따르고 서로 독립이다)

① $\hat{\beta}_0$은 β_0의 불편추정량이다.
② $\hat{\beta}_1$은 β_1의 불편추정량이다.
③ $\hat{\beta}_0$의 분산은 $Var(\hat{\beta}_0)=\sigma^2\times\left(\dfrac{1}{n}+\dfrac{(\overline{x})^2}{\sum_{i=1}^{n}(x_i-\overline{x})^2}\right)$이다.
④ $\hat{\beta}_1$의 분산은 $Var(\hat{\beta}_1)=\dfrac{\sigma^2}{\sum_{i=1}^{n}(x_i-\overline{x})^2}$이다.

문 22. 공정 1, 2, 3에서 생산된 제품의 강도에 차이가 있는지를 알아보기 위해 세 공정에서 생산된 제품을 임의로 선택하고 강도를 측정하여 요약한 결과가 다음과 같을 때, 이를 이용한 분산분석의 설명으로 옳지 않은 것은? (단, n_i는 공정 i에서 선택된 제품 수, y_{ij}는 공정 i에서 j번째 관측 강도, $\overline{y}_{i\cdot}$는 공정 i에서의 평균 강도, $\overline{y}$는 강도의 전체 평균을 나타낸다)

공정(i)	1	2	3
n_i	4	5	6
$\overline{y}_{i\cdot}$	117	120	122
$\sum_{j=1}^{n_i}(y_{ij}-\overline{y}_{i\cdot})^2$	60	80	100
$\overline{y}$	120		
$\sum_{i=1}^{3}\sum_{j=1}^{n_i}(y_{ij}-\overline{y})^2$	300		

① 총제곱합의 자유도는 14이다.
② 오차제곱합은 240이다.
③ 공정에 대한 평균제곱은 20이다.
④ 귀무가설하에서 검정통계량의 분포는 자유도가 $(2, 12)$인 $F-$분포이다.

문 23. 다음은 납축전지의 효율이 온도와 습도 요인에 따라 영향을 받는지를 알아보기 위한 실험의 분석 결과를 정리한 분산분석표의 일부이다. 실험은 3가지 온도와 4가지 습도의 모든 수준 조합에서 한 번씩 이루어졌다. 이에 대한 설명으로 옳은 것은?

요인	제곱합	자유도	$F-$값	$p-$값
온도	42			0.027
습도	18			0.216
오차	18			

① 오차항의 자유도는 8이다.
② 습도에 대한 $F-$값은 4이다.
③ 유의수준 5 %에서 납축전지의 효율이 온도의 수준에 따라 모두 같다고 할 수 있다.
④ F가 자유도 $(2, 6)$인 $F-$분포를 따르는 확률변수일 때, 온도에 대한 $p-$값은 $P(F>7)$이다.

문 24. 모수 θ에 따른 이산형 확률변수 X의 분포 $f(x;\theta)$는 다음과 같다.

$f(x;\theta)$ \ θ	$\theta=0$	$\theta=2$	$\theta=4$	$\theta=6$
$f(1;\theta)$	$\frac{1}{3}$	$\frac{1}{2}$	0	$\frac{1}{5}$
$f(2;\theta)$	$\frac{1}{3}$	0	$\frac{1}{4}$	$\frac{2}{5}$
$f(3;\theta)$	$\frac{1}{3}$	$\frac{1}{2}$	$\frac{3}{4}$	$\frac{2}{5}$

위 분포로부터 크기 2인 확률표본 X_1과 X_2가 각각 1과 3으로 관측되었을 때, θ에 대한 최대가능도추정치(maximum likelihood estimate)는?

① 0
② 2
③ 4
④ 6

문 25. 두 연속형 확률변수 X와 Y의 결합확률밀도함수가 다음과 같을 때, $P(X+Y\le 1)$의 값은?

$$f(x,y)=\begin{cases}\frac{1}{2}, & 0\le x\le 1,\ 0\le y\le 2\\ 0, & \text{그 외}\end{cases}$$

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